Gene discovery in developmental neuropsychiatric disorders: Clues from chromosomal rearrangements

Thomas V. Fernandez

YALE UNIVERSITY

2005

Gene discovery in developmental neuropsychiatric disorders: Clues from chromosomal rearrangements

A Thesis Submitted to the
Yale University School of Medicine
In Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

by
Thomas V. Fernandez
2005

Abstract

GENE DISCOVERY IN DEVELOPMENTAL NEUROPSYCHIATRIC DISORDERS: CLUES FROM CHROMOSOMAL REARRANGEMENTS. Thomas Fernandez[1], Thomas Morgan[2], Nicole Davis[1], Ami Klin[1], and Matthew W. State[1,2,3]. [1]Child Study Center, [2]Department of Genetics, and [3]Yale Center for Human Genetics and Genomics, Yale University, School of Medicine, New Haven, CT.

The molecular characterization of rare affected individuals presenting with chromosomal abnormalities has made recent important contributions to disease gene identification in developmental disorders and autism. Here we review the success of this approach compared to more traditional approaches toward disease gene identification for complex developmental neuropsychiatric disorders. We report on cases of children affected by autism and pervasive developmental disorder (PDD) with different balanced chromosomal translocations. By using fluorescent *in situ* hybridization (FISH) to fine-map the breakpoints of these chromosomal rearrangements, we discovered the physical disruption of one gene transcript (*Contactin4, CNTN4*) in a boy with PDD and a phenotype consistent with 3p deletion syndrome. Our results demonstrate the association of *CNTN4* disruption with the 3p deletion syndrome phenotype, strongly suggest a causal relationship, and point to an important role for *CNTN4* in normal and abnormal CNS development. We subsequently screened a cohort of patients (N=97) with autism and autism spectrum disorders (ASDs) and normal karyotypes, along with control subjects (N=234), for mutations in coding regions of this gene. One missense mutation was identified in a multiplex ASD family and not found in controls, although this mutation did not completely co-segregate with affected status. Finally, by mapping the breakpoints of an inherited translocation in three autistic siblings, we identified an interesting candidate gene for autism susceptibility on chromosome 11 that warrants additional study.

Acknowledgements

I would like to thank my advisor, Matthew State, MD, PhD, for his teaching, encouragement, and support during the time I have spent in his lab over the past year and a half. I am also grateful to Nicole Davis and the many other members of the State Lab, who took time to show me the ropes and are always there to lend a hand. Finally, I extend my deepest gratitude to the patients and their families for their participation in this research.

This work was supported by a Career Development Award NCRR-K23 RR 16118-02; a Pilot Grant for Translational and Interdisciplinary Research from the Yale School of Medicine and grants from the Shepherd Foundation (all to MWS). Stipend support was provided through Yale University School of Medicine Office of Student Research Short Term Research Funding (2003-2005, to TVF).

TABLE OF CONTENTS

INTRODUCTION

Evidence for the heritable nature of developmental neuropsychiatric disorders such as autism has been known for several decades. Twin and family studies from the 1970s have clearly established the importance of genetic factors in autism and autism spectrum disorders (ASDs), with monozygotic twins showing concordance rates of 40-60% and 70-90%, respectively, compared with rates of 0-25% in dizygotic twins (1-3). The sibling recurrence risk for autism is estimated at 2-6%, which is 50-100 times greater than the population prevalence of 0.04-0.1% (3-5). With an estimated heritability of approximately 90%, the rate by which the incidence of autism and ASDs drops among successive generations of relatives suggests that disease susceptibility arises from a complex mode of inheritance, involving the combined effects of multiple, possibly interacting genes (6, 7).

Over the last decade, different approaches have been utilized in an attempt to identify genes involved in the etiology of complex developmental disorders. The basic principles behind these strategies are briefly reviewed below.

Traditional approaches to disease gene identification

The two most widely employed approaches to disease gene identification, or mapping, are linkage and association studies. Linkage is based on identifying markers that co-segregate with the disease within families. This approach classically uses multigenerational pedigrees containing several affected individuals in which it is possible

to infer the mode of inheritance of the disease locus and specify the relevant parameters. In genetically complex disorders such as autism, the mode of inheritance is generally unknown and large affected pedigrees are unavailable. Therefore, alternative "model free", or non-parametric, linkage methods are used that do not require specification of an inheritance model but instead rely on assessing allele sharing between two or more affected individuals within multiple nuclear families. The goal is to identify genetic markers at which affected pairs share alleles more often than expected by chance, as such markers may be linked to a susceptibility gene. Although these non-parametric linkage methods are robust, they have relatively low power to detect genes of small effect; in this scenario, association studies are more powerful.

Association strategies are based on identifying alleles that occur at significantly different frequencies in samples of affected versus control individuals. Case-control studies utilize unrelated controls, and careful matching is required to avoid false positive results due to cases and controls being drawn from different gene pools. Family-based association designs, such as the transmission disequilibrium test (TDT), overcome this potential difficulty by using the frequencies of parental alleles that are not transmitted to affected offspring as internal controls (8). Association of a particular allele with a disease can occur if the allele is implicated in the etiology of the disease, or if it is in linkage disequilibrium with the susceptibility locus. Since association is usually only detectable if the allele is very close to or within the susceptibility gene, this approach is predominantly used to test polymorphisms within known candidate genes, or as a tool for fine mapping a region of interest identified by linkage studies (9).

Outcome of traditional approaches in psychiatric genetics

Overall, the field of psychiatric genetics has had few successes in attempts to find relevant genes through standard linkage analysis. While several genome-wide linkage studies of autism and ASDs have suggested multiple intervals that may contain relevant loci, this approach has not yet led to the identification of a specific transcript conferring disease risk. The most recent and largest of ten independent genome-wide scans of patients with ASDs found no statistically significant evidence for linkage, but showed 'suggestive' evidence for linkage on chromosomes 17q, 5p, 11p, 4q, and 8q, with multipoint maximum LOD scores between 1.60-2.83 (10). It is likely that a combination of genetic heterogeneity, clinical heterogeneity, an uncertain mode of inheritance, and the interaction of multiple genes that increase *susceptibility* to autism, rather than directly cause it, have hindered efforts at disease gene identification. As a result of these factors, the correlation between disease status and any given DNA marker allele in a genome-wide linkage study tends to be relatively weak, and the implicated regions, or linkage peaks, typically encompass a large and often unmanageable number of positional candidate genes, the vast majority of which are unrelated to the disorder.

An alternate approach to disease gene identification

Due to the difficulties with traditional approaches for disease gene identification outlined above, alternative approaches are warranted. One such strategy for candidate gene discovery relies on the characterization of rare affected individuals who present with chromosomal abnormalities ("genetic outliers"). Mapping the breakpoints in chromosomal deletions and translocations associated with a disorder can be particularly relevant in the attempt to pinpoint candidate gene loci, as this may clearly identify a disrupted gene or genes. This approach has been used successfully in helping to identify the genes for multiple disorders, including Duchenne muscular dystrophy, familial Alzheimer's disease (11) and a host of mental retardation syndromes (12).

In psychiatry, this approach of studying genetic outliers has recently led to the identification of *Neuroligin 4* (*NLGN4*) as one of the first transcripts corresponding to "idiopathic autism". In this case, cytogenetic studies found X-chromosome deletions in three autistic patients (13). This led to screening of a candidate interval, resulting in the demonstration of a functional mutation that segregated with ASDs in a European family (14). Two subsequent independent human genetic studies have confirmed these initial findings for *NLGN4* (15, 16) and another has demonstrated the functional relevance of the identified *NLGN4* mutations (17).

These recent results with autism demonstrate the opportunities presented by the study of unusual patients with cytogenetic abnormalities. Despite this opportunity, chromosomal anomalies remain understudied in developmental disorders such as ASDs (18). While such efforts are most likely to identify genes of major effect that are of

consequence for only a small sub-group of individuals with ASDs, they promise to yield important insights into physiologic pathways and genetic mechanisms involved in more common forms of these disorders.

Statement of purpose and hypothesis

Using molecular genetic techniques to investigate and characterize the breakpoints of chromosomal abnormalities in individual patients with developmental delay and ASDs will identify interesting candidate genes involved in their pathology. Based on these findings, screening genomic DNA for sequence variations in the identified gene(s) in a larger cohort of affected and control subjects without known cytogenetic abnormalities will suggest whether such regions may be involved in only rare or more common cases of these disorders. The identification of such mutations in even a small number of cases may provide insights into underlying disease mechanisms.

MATERIALS AND METHODS

Analysis of individual patients with known chromosomal abnormalities

Patient assessment

Molecular analysis of four patients with ASDs and known cytogenetic abnormalities was completed using the methods described below. These patients were obtained through referral to the Yale Child Study Center (YCSC) and the Yale Outpatient

Genetics Clinic. Clinical assessment of these patients was performed by members of the faculty and staff of YCSC and the Department of Genetics. Physical examinations, medical and family histories, and scores on standardized ASD rating scales (described below) were obtained for each patient.

Karyotype analysis

Standard chromosome preparations from all participants were made from PHA-stimulated peripheral blood lymphocytes. Conventional G-banded chromosome analyses were performed on metaphase cells with a minimum resolution level of 550 bands.

Fluorescence in situ hybridization analysis

In order to fine map chromosomal rearrangements detected by cytogenetics, fluorescent *in situ* hybridizations (FISH) of metaphase nuclei were carried out as previously described (19). Bacterial Artificial Chromosome (BAC) clones corresponding to the regions of interest identified on G-banding analysis were identified using the UCSC Human Genome Browser at http://genome.ucsc.edu/ (July 2003 freeze) and obtained from an RPCI-11 (Roswell Park Cancer Institute) BAC library. These BAC clones were cultured, the human DNA inserts were extracted, and then fluorescently labeled by nick translation. Labeled BAC DNA was ethanol precipitated along with Cot1 DNA to block repetitive sequences. Each clone was then used as a probe and hybridized to metaphase spreads consisting of the patient's lymphocytes. Following hybridization, the nuclei were counterstained with DAPI to distinguish nuclei from background and observed under multi-wavelength fluorescent microscopy.

Array-based copy number analysis

The GeneChip Mapping 100K Set (Affymetrix, Inc., Santa Clara, CA) was used to examine the chromosomes of several patients in this study. The 100K Set is composed of two arrays, each with greater than 50,000 SNPs with 8.5 kb median physical distance and 23.6 kb average distance between SNPs. SNP GeneChip assay was performed according to the manufacturer's manual. Total genomic DNA (250ng) was digested with restriction enzymes XbaI and HindIII (New England Biolabs) for 2 hours at 37^0C to reduce the complexity of the genome and then, ligated to an adapter that recognizes the cohesive four base pair (bp) overhangs using T4 DNA ligase at 16^0C for 2.5 h. Subsequently, adapter ligated DNA fragments were amplified using a generic primer that recognizes the adapter sequence. The PCR product sizes ranged from 250 to 2,000 bp. A total of 40 μg of purified PCR products were fragmented with DnaseI. Fragmented PCR products were biotin- labeled and hybridized to GeneChip® Human Mapping 50K Array Xba240 and GeneChip® Human Mapping 50K Array Hind 240 for 16 h at 45 ^{0}C in an Affymetrix GeneChip Hybridization oven. After hybridization, the arrays were washed, stained, and scanned using an Affymetrix Fluidics Station F450, and images were obtained by using Affymetrix GeneChip scanner 3000. We analyzed cell intensity data from the Affymetrix® GeneChip® Human Mapping 100K Set using GeneChip DNA Analysis Software Version 3.0.2.8 (GDAS). We then used the Affymetrix® GeneChip® Chromosome Copy Number Tool (CCNT) to search genome-wide chromosomal gains and looses at high resolution. This tool uses a reference set of ≥110 ethnically diverse normal individuals to compare with patient samples by generating averaged p-value.

Thresholds for copy number gain and loss (≤1.4 regarded as deletion and ≥2.8 considered as gain) is detected by using two known samples with deletion and amplification. SNP annotations were obtained from NetAffx™ Analysis Center.

Candidate gene identification

After narrowing the breakpoints to the smallest region possible using FISH with labelled BAC probes, known genes, full length mRNAs, and spliced expressed sequence tags (ESTs) in the breakpoint regions were surveyed using the UCSC Human Genome Browser at http://genome.ucsc.edu/ (July 2003 freeze).

Screening candidate genes in a larger population of affected patients

Patient recruitment

Based on the findings from cytogenetic analyses of individual patients with ASDs, reported here, a candidate gene was screened for mutations in a cohort of ASD patients without known cytogenetic abnormalities. Clinical data and DNA were obtained from the Autism Genetics Research Exchange (AGRE, http://www.agre.org), a DNA repository and family registry, housing a publicly available database of genotypic and extensively characterized phenotypic information. AGRE was founded by the Cure Autism Now Foundation (http://www.canfoundation.org) and is now supported by the National Institute of Mental Health (NIMH) as part of the NIMH genetics initiative. Families are recruited through a variety of methods (e.g. physician referral, website contact, and family meetings and seminars). They are ascertained on the basis that at least

two family members met criteria for a diagnosis of an ASD - autism, Asperger's syndrome, or pervasive developmental disorder (PDD). To date, there are a total of 540 AGRE families available for distribution, 490 of which are multiplex.

AGRE population assessment

The diagnosis of autism has been made using the standard Autism Diagnostic Interview–Revised (ADI-R) algorithm (20), with further phenotypic characterization using the Autism Diagnostic Observational Schedule-Generic (ADOS-G) (21), Vineland Adaptive Behavior Scale (22), and other cognitive tests. In addition to the diagnosis of autistic disorder, AGRE provides two other diagnostic classifications ('Not Quite Autism' and 'Broad Spectrum') to characterize those individuals on the broader autism spectrum who do not meet full ADI-R criteria for autism. All ADI and ADOS raters undergo ongoing reliability checks to prevent any drift in diagnosis (20, 21). In addition, pediatric neurologists assess photographic dysmorphology, perform physical and neurological examinations, and collect medical and family history. Probands with possible non-idiopathic autism resulting from perinatal trauma, from an identified genetic syndrome (e.g., fragile X syndrome), or from other medical causes are noted, although this represents only a small percentage of cases (23); these probands were excluded from the current study.

ADI-R

Diagnosis of autism is established by the well-validated Autism Diagnostic Interview Revised (ADI-R) diagnostic algorithm (20). The ADI-R is currently the gold

standard for research diagnosis and is based on the classifications of the International Statistical Classification of Diseases and Related Health Problems, 10th revision (24) as well as the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) (25). The interview contains 111 items and focuses on behaviors in three content areas: (1) quality of social interaction, (e.g., emotional sharing, offering and seeking comfort, social smiling and responding to other children); (2) communication and language (e.g. stereotyped utterances, pronoun reversal, social usage of language); and (3) repetitive, restricted, and stereotyped interests and behavior (e.g. unusual preoccupations, hand and finger mannerisms, unusual sensory interests). Responses are scored by the clinician based on the caregiver's description of the child's behavior. To meet criteria for autism, a child must meet criteria in each of the three content areas as well as exhibit some abnormality in at least one content area by 36 months of age. The cutoff total score for the communication and language content area is 8 for verbal subjects and 7 for nonverbal subjects. For all subjects, the cutoff for the social interaction content area is 10, and the cutoff for restricted and repetitive behaviors is 3. The instrument has been shown to be reliable and to successfully differentiate young children with autism from those with mental retardation and language impairments.

ADOS-G

The Autism Diagnostic Observation Schedule- Generic (ADOS-G) is a semi-structured assessment of communication, social interaction and play or imaginative use of materials for individuals suspected of having autism or other pervasive developmental disorders (PDD) (21). It is a combination of two earlier instruments: the Autism

Diagnostic Observation Schedule (ADOS) (26), a schedule intended for adults and children with language skills at a minimum of the three-year-old level, and the Pre-Linguistic Autism Diagnostic Observation Scale (PL-ADOS) (27), a schedule intended for children with limited or no language, as well as additional items developed for verbally fluent, high-functioning adolescents and adults. The ADOS-G consists of four modules, each of which is appropriate for children and adults of differing developmental and language levels, ranging from no expressive or receptive language to verbally fluent adults.

The ADOS-G consists of standard activities that allow the examiner to observe the occurrence or non-occurrence of behaviors that have been identified as important to the diagnosis of autism and other pervasive developmental disorders across developmental levels and chronological ages. Overall ratings can then be used to formulate a diagnosis through the use of a diagnostic algorithm for each module. For purposes of diagnosis, use of this instrument should be accompanied by information from other sources, particularly a detailed history from parents whenever possible (20).

DNA mutation detection

One relatively new method used in this study to screen gene sequences rapidly for the presence of sequence variations is denaturing high-performance liquid chromatography (dHPLC). Initially described by Oefner and colleagues (28, 29), dHPLC has become a reproducible and robust high-throughput method for discovery of novel single nucleotide polymorphisms (SNPs) as well as small deletions or insertions with sensitivity comparable to that of direct DNA sequencing (30, 31) and superior to that of

alternative gel based analysis techniques (i.e. SSCP) (32). The principle behind the dHPLC method relies on detecting heteroduplex DNAs in PCR products amplified from genomic DNA in an individual that is heterozygous for a sequence variant. The sensitivity of heteroduplex detection relies on maintaining the dHPLC column at a specific temperature, which favors the increased thermodynamic instability of a PCR product with a single-base mismatch (i.e. a heteroduplex) (33, 34) over that of a PCR product containing perfectly matched base pairing (i.e. a homoduplex). DHPLC has been shown to clearly resolve mutations in various genes with detection rates ranging from 92.5% to 100% (32, 35).

Amplification of samples

PCR products prepared from AGRE patient genomic DNA served as the DNA fragments analyzed for the presence of mutations. Primers were designed to amplify at least 15 bp flanking each intron/exon boundary for all coding regions of the candidate gene. These primer pairs were designed by examining annotated genomic sequence databases (http://genome.ucsc.edu and http://www.ncbi.nlm.nih.gov) and using Primer3, an online conventional primer design software package (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Primers and sequences are listed in Appendix Table A1. Primers were produced by Invitrogen Corporation (Carlsbad, CA).

Primers were optimized to one of six different annealing temperatures (55°C, 56°C, 56.8°C, 58.1°C, 59.1°C, 60°C). Optimum annealing temperatures for each primer pair are listed in Appendix Table A1.

PCR conditions for the various primer pairs were as follows. Each reaction was performed in a total volume of 30 µl and contained 1 µl of DNA at 50 ng/µl, 15 µl of MasterAmp™ 2x PCR PreMix D Buffer (Epicentre Technologies, Madison, WI), 1 µl of 10 nM forward primer, 1 µl of 10 nM reverse primer, and 0.125 µL of 5U/µl Taq DNA Polymerase (Roche Diagnostics Corporation, Indianapolis, IN). The PCR temperature program consisted of a 4 min hot start at 94°C followed by 30 cycles of 30 sec of denaturation at 94°C, 30 sec at the optimal annealing temperature (see Appendix Table A1), and 45 sec at 72°C for PCR product elongation. There was a final 10 min elongation step at 72°C, and then the reaction was stored at 4°C until ready to be analyzed by dHPLC.

Variant screening using dHPLC

Following PCR amplification, the samples were denatured at 95°C for 10 min and then cooled to 30°C over a period of 10 min to enhance the formation of DNA heteroduplexes. Samples were then run on a Transgenomic WAVE® dHPLC system with the DNASep® dHPLC column (Transgenomic, San Jose, CA). The WAVE system consists of a cooled 96-well autosampler, column oven, pumps, buffer degasser, variable wavelength UV detector, a 20-µl sample loop, and a PC-based data collection system. Chromatography on the WAVE system is performed with a dual buffer system, where the buffers used were as follows: Buffer A, 0.1 M triethylammonium acetate (TEAA, Transgenomic, San Jose, CA), pH 7.4, and Buffer B, 0.1 M TEAA and 25% acetonitrile (CAN, Transgenomic, San Jose, CA). Both the loading buffer and syringe wash solution consisted of 10% ACN for the DNASep column. In addition, the Transgenomic WAVE

system injects 200 µl of a 75% ACN column wash solution, of which 200 µl was injected after each run. TEAA, ACN, and water used for the preparation of buffers were of HPLC grade.

Using this system, 7 µL of PCR product was loaded onto the preheated column and was eluted by an increasing acetonitrile gradient at a certain partially denaturing temperature which was specific to each PCR amplicon. All dHPLC conditions, including melting temperatures and buffer gradients specific to each PCR amplicon, were determined using the DHPLC Melt Program at http://insertion.stanford.edu/melt.html. Some amplicon fragments required dHPLC analysis at more than one temperature in order to detect possible mutations along the entire fragment (all DHPLC temperatures are listed in Appendix Table A1). Eluent flow rates for all analyses were 0.9 ml/min. Eluted DNA fragments were detected by UV absorption at a wavelength of 260 nm and were evaluated for abnormalities in elution profile as described below.

Mutation Calling

Sequence variations are identified based on a comparison between the chromatogram waveforms of wild type amplicons (i.e. no sequence variations) and chromatograms of amplicons carrying heterozygous mutations, recorded under identical dHPLC conditions. A characteristic example of mutation calling is illustrated in Figure 1.

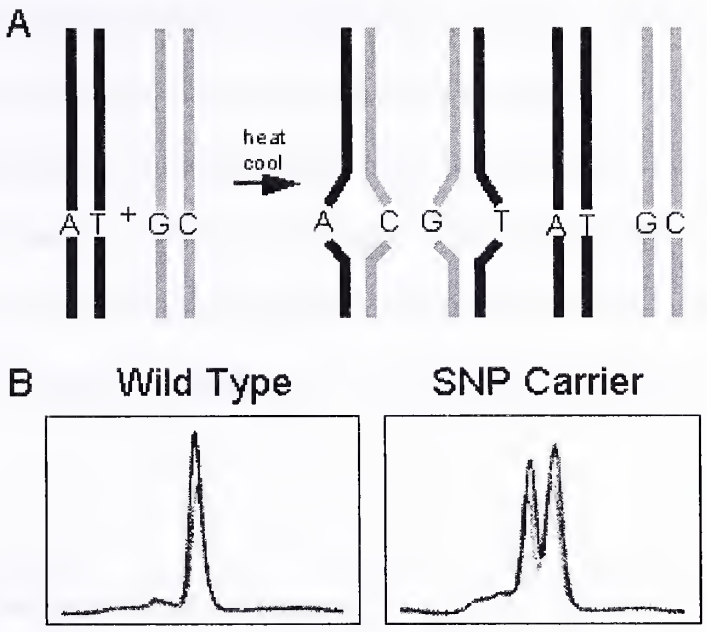

Figure 1. Duplex formation and dHPLC mutation screening. A. Mixtures of wild type (black) and mutant (grey) amplicons, which are either co-amplified during the PCR or generated by post-PCR mixing, are denatured by heating and reannealed by slow cooling, leading to the formation of four distinct species, two original homoduplexes and two heteroduplexes. Sequence variations resulting in heteroduplex formation include SNP as well as short insertions and deletions. **B.** Wild type and mutant homoduplexes melt at higher temperatures than mismatch containing wild type/mutant heteroduplexes. Differences in melting temperatures are the basis for the identification of mutations by dHPLC, which will display different waveforms based on different melting temperatures.

The wild type chromatogram exhibits a typical single-peak pattern, because the amplified wild type sample contains a single homoduplex amplicon. With increasing analysis temperature, the PCR product progressively melts, resulting in a steady increase in its single-stranded character and causing a steady reduction in retention time on the DNASep column. This is the result of decreased affinity of single-stranded DNA to the column matrix as compared to double-stranded DNA. The single peak chromatogram of a wild type homoduplex serves as a reference for mutation calling.

The presence of sequence variations (polymorphisms or mutations) in an amplicon leads to the formation of four distinct duplexes (Figure 1A) during the

denaturation and reannealing step preceding dHPLC screening. This presents on a chromatogram as the appearance of additional peaks (Figure 1B). Because homoduplexes have higher melting temperatures than the corresponding mismatch-containing heteroduplexes, the retention time of peaks representing homoduplexes is always longer than the retention times of their heteroduplex counterparts under identical gradient conditions and at analysis temperatures at or near the melting temperature of the homoduplex.

DNA sequencing of putative variants

Those individual PCR samples that appeared to display a heteroduplex-type peak were sequenced to confirm the presence of a heterozygous DNA variant—either point mutation, SNP, insertion, or deletion. PCR products from the original reaction were purified and sequenced by Genaissance, Inc. (New Haven, CT).

Sequencing data was viewed using the Sequencher software package (Gene Codes, Inc., Ann Arbor, MI) and each sequence was matched to the corresponding chromatogram waveform in order to confirm the presence of a DNA variant. For each sample in which a DNA variant was confirmed, a translated amino acid sequence was matched to a reference (wild type) amino acid sequence in order to determine whether the DNA variation would result in an amino acid change.

Statement regarding level of involvement

I conducted all elements of this project, with the exception of patient recruitment, patient clinical evaluations, and karyotype analysis. As mentioned above, members of the clinical and research faculty and staff at the Yale Child Study Center and Department of Genetics recruited and evaluated the individual cases with known chromosomal abnormalities. Patient recruitment and clinical assessment of AGRE patients were completed by the AGRE and Cure Autism Now Foundation research teams (Los Angeles, CA). Karyotype analysis was performed by Nicole Davis, a cytogeneticist in the Department of Genetics and Child Study Center.

RESULTS

Case 1 – PDD and a balanced (3;10) translocation

Clinical assessment

CT is a boy aged 7 years 11 mo with borderline intellectual functioning, distinctive facial (Figure 2) and other physical features (Appendix Table A2), and behavioral problems who was seen in the Genetics Outpatient Clinic at Yale and evaluated under a research protocol approved by the Yale Human Investigation Committee with informed consent. The patient was the product of a non-consanguineous union, delivered at 40 weeks gestation from his 29 year-old primiparous mother by cesarean section performed because of failure to progress. There was no history of alcohol consumption or medication use during pregnancy and no known exposure to

teratogens. He had a difficult facial presentation, with mild perinatal respiratory depression and an Apgar score of 3 at 1 min. Cord blood pH was 7.28. After brief positive pressure ventilation, his Apgar score was 9 at 5 min. Birth weight was 3,787 g and length was 19 inches. Cytogenetic analysis was performed shortly after birth, which revealed a balanced translocation [46, XY,t(3;10)(p26;q26)]. Both parents had normal karyotypes. No other family members were reported to have mental retardation or dysmorphic features.

The child's early motor development was unremarkable: he rolled over at 4 months and was sitting with support at 6.5 months. At 7 months, he was able to stand with assistance, and he walked at 14 months. His development of language, in contrast, was markedly delayed, with the absence of distinct single words until 3.5 years. CT underwent a cognitive assessment at age 7 years 2 months and was found to have a full scale IQ of 73 on the Wechsler Intelligence Scale for Children, Third Edition (WISC-III) (Mean = 100; SD = 15), which places his cognitive skills within the borderline range of intellectual functioning. He obtained a verbal IQ of 81 and a performance (or nonverbal) IQ of 70. Significant weaknesses in his profile related to attention to visual detail (subtest score of 3; mean = 10, SD = 3), visual-spatial reconstruction skills (subtest score of 3), and understanding of social mores and expectations (subtest score of 3). These deficits contrasted with his significant strength in the area of fund of knowledge (subtest score of 12).

Figure 2. Case 1. Eight-year-old boy with developmental delay (PDD-NOS) and dysmorphic facial features. The child has had surgical correction of prominent ptosis and a notable low frontal hairline. (Photographs used with consent).

Assessment of adaptive functioning at YCSC using the Vineland Adaptive Behavior Scales (22) showed impairments in all four domains of adaptive behavior covered by this test. His standard scores and percentile ranks (%) were as follows: communication 8 (14%), daily living skills 50 (<0.1%) socialization 66 (1%), and motor skills 76 (5%). His adaptive composite score, which summarizes performance across all of these domains, was 61 (0.5%).

CT was also evaluated for autistic disorder using the Autism Diagnostic Interview-Revised (ADI-R) (20) and Autism Diagnostic Observation Schedule- Generic (ADOS-G) (21). ADI-R scores were as follows: communication 13 (threshold score for autism is 8), social interaction 25 (threshold score for autism is 10), stereotyped behavior 3 (threshold score for autism is 3). ADOS-G scores were as follows: communication domain 2 (threshold score for autism is 3, for PDD-NOS is 2), social domain 4 (threshold

score for autism is 6, for PDD-NOS is 4), imagination/creativity 1 (no threshold scores proposed), stereotyped behaviors and restricted interests 1 (no threshold scores proposed).

CT's past medical history was unremarkable. He was noted to be in excellent physical health and had required surgery only for correction of ptosis. He had been followed closely from a very young age by his pediatrician as a consequence of his dysmorphic features and known cytogenetic rearrangement and was not found to have any evidence of GI, cardiac, or renal abnormalities on physical exam or by history. A referral to a pediatric neurologist for behavioral issues at age 6 years revealed a normal neurological exam, except for the identification of mild hypotonia. Neither neuroimaging nor electroencephalogram was deemed to be clinically indicated.

Physical examination at age 7 years 11 months revealed a height of 115.6 cm (5^{th}-10^{th} percentile), a weight of 21.4 kg (5^{th} percentile), and a head circumference of 52 cm (50^{th}-75^{th} percentile). The patient showed high myopia and astigmatism, a low frontal hairline with right upsweep, forehead hypertrichosis, and bushy eyebrows without frank synophrys. He was also noted to have surgically improved ptosis and slight epicanthal folds, a broad nasal bridge with mild hypertelorism (measured at +1.5 SD at age 21 mo), downslanting palpebral fissures, downturned corners of the mouth, prominent but normal-sized ears with superior helices sagging downward, slight fifth-finger clinodactyly, branched transverse palmar crease of right hand, mild bilateral syndactyly of the second and third toes, hypertrichosis of his thighs and legs, and mild hypotonia. Deep tendon reflexes were normal.

Cytogenetic analysis

Molecular studies were undertaken to fine map the (3;10) translocation using FISH. Several probes mapping centromeric and telomeric to the translocation breakpoints on both chromosomes were identified (Table 1). Subsequent to the identification of BACs mapping on both sides of the translocation breakpoint on each chromosome, BACs within the translocation intervals were used, eventually identifying probes spanning both the chromosome 3 and 10 breakpoints (Figures 3 and 4). Two BACs spanning the chromosome 3 breakpoint, RPCI-11 299N33 and RPCI-11 121K9, were found to map in their entirety to the first intron of the gene *Contactin 4* (*CNTN4* [MIM 607280]) (Figure 4A), establishing that the translocation breakpoint disrupted the gene within the 5' untranslated region (UTR) of the mature *CNTN4* mRNA.

Table 1. Case 1: BAC probes investigated using fluorescence *in situ* hybridization[A]

BAC ID	Base position (July 2003 Freeze)	Chromosomal band	Location relative to breakpoint
RP11-1082A18	chr3:10,298,689-10,516,466	3p25.3	centromeric
RP11-572M14	chr3:10,011,785-10,180,797	3p25.3	centromeric
RP11-1020A11	chr3:9,778,862-9,995,772	3p25.3	centromeric
RP11-19E8	chr3:8,917,083-9,067,340	3p25.3	centromeric
RP11-128A5	chr3:8,686,926-8,857,271	3p25.3	centromeric
RP11-34L16	chr3:8,442,935-8,594,857	3p25.3	centromeric
RP11-470E10	chr3:7,841,368-8,025,425	3p26.1	centromeric
RP11-507D6	chr3:6,997,984-7,199,514	3p26.1	centromeric
RP11-318I14	chr3:6,283,022-6,437,405	3p26.1	centromeric
RP11-161L3	chr3:5,744,835-5,909,633	3p26.1	centromeric
RP11-129J10	chr3:5,027,803-5,180,950	3p26.1	centromeric
RP11-453A3	chr3:4,158,926-4,329,972	3p26.1	centromeric
RP11-245A6	chr3:3,545,257-3,707,298	3p26.2	centromeric
RP11-198P17	chr3:2,928,181-3,117,274	3p26.2	centromeric
RP11-94M9	chr3:2,804,247-2,975,122	3p26.2	centromeric
RP11-785A7	chr3:2,690,183-2,862,993	3p26.2-3p26.3	centromeric
RP11-82O3	chr3:2,566,282-2,742,197	3p26.3	centromeric
RP11-63O1	chr3:2,443,361-2,617,387	3p26.3	centromeric
RP11-211K13	chr3:2,307,996-2,470,413	3p26.3	centromeric
RP11-762O12	chr3:2,287,581-2,441,071	3p26.3	centromeric
RP11-299N3	chr3:2,181,279-2,367,269	3p26.3	spanning
RP11-129K1	chr3:2,132,133-2,290,752	3p26.3	spanning
RP11-176P14	chr3:2,021,987-2,193,951	3p26.3	telomeric
RP11-416N8	chr3:1,244,095-1,420,317	3p26.3	telomeric
RP11-392M7	chr3:1,064,436-1,252,749	3p26.3	telomeric
RP11-114K9	chr3:333,643-518,322	3p26.3	telomeric
RP11-306H5	chr3:159,323-343,406	3p26.3	telomeric
RP11-338L11	chr10:116,448,883-116,645,816	10q25.3	centromeric
RP11-169K19	chr10:117,157,040-117,301,164	10q25.3	centromeric
RP11-96N16	chr10:117,471,461-117,637,580	10q25.3	centromeric
RP11-295023	chr10:117,733,868-117,949,898	10q25.3	centromeric
RP11-498B4	chr10:118,011,104-118,209,364	10q25.3	centromeric
RP11-5G18	chr10:118,364,598-118,528,414	10q25.3-10q26.11	centromeric
RP11-328K15	chr10:119,130,080-119,172,006	10q26.11	centromeric
RP11-355F22	chr10:119,170,007-119,301,205	10q26.11	spanning
RP11-136I7	chr10:119,181,243-119,340,828	10q26.11	spanning
RP11-280F22	chr10:119,299,202-119,400,731	10q26.11	telomeric
RP11-3H12	chr10:119,316,263-119,394,906	10q26.11	telomeric
RP11-354M20	chr10:119,398,732-119,592,400	10q26.11	telomeric

[A] BAC probes used for FISH analysis: The RPCI-11 identifiers, chromosome number and band, location in base pairs (July 2003 freeze), and position determined relative to the breakpoint(s) are shown for all BACs hybridized. BAC probes noted as centromeric and telomeric yielded two hybridization signals on metaphase spreads. Spanning probes resulted in three hybridization signals

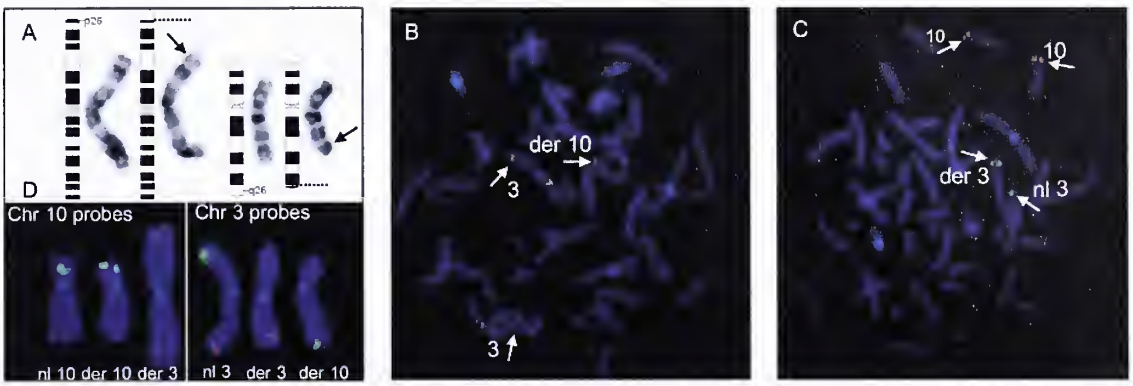

Figure 3. Cytogenetic and molecular cytogenetic studies of a *de novo* (3;10) translocation (Case 1). A. G-banded karyotype of the patient's lymphocytes showing a t(3p26;10q26) translocation. Arrows highlight the breakpoints on the respective chromosomes, and dotted lines show the corresponding position on the ideograms. 10/10 metaphase spreads examined showed the identical translocation. **B.** BAC probe 299N3 (*red*) spanning the chromosome 3p breakpoint. Hybridization signals are seen on both normal and derivative chromosomes 3. A third red hybridization signal is also seen on the derivative chromosome 10, as a result of the breakpoint transecting the region corresponding to the experimental probe. A control probe (*green*) maps to the q arm of chromosome 3. **C.** Probe 355F22 (*red*) spanning the chromosome 10q breakpoint. Hybridization signals are present on the normal and derivative chromosomes 10. A third red signal is seen on the telomeric region of chromosome 3. Control probes (*green*) map to the short arm of chromosome 3. **D.** Subtelomeric FISH results are shown. Chromosome 10 probes are in the left box—the normal (nl) chromosome 10 with both green and red probes, the derivative (der) 10 with only a green probe, and the derivative chromosome 3 with only the red probe are shown. The box on the right shows chromosome 3 probes, with the normal chromosome showing both red and green probes, the derivative 3 showing only the red probe, and the derivative 10 showing only the green probe.

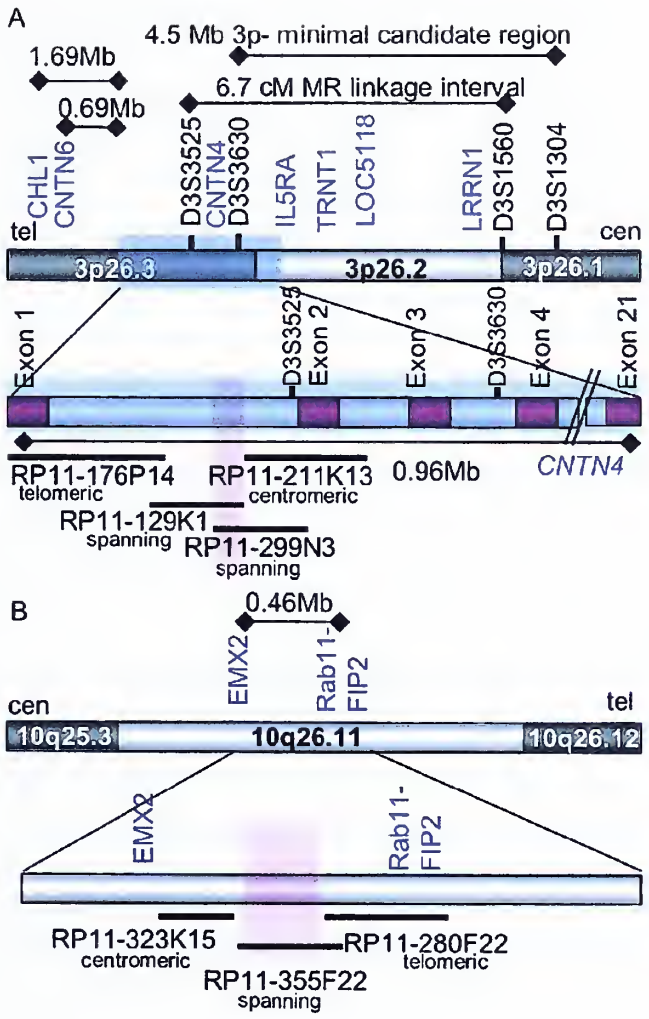

Figure 4. Diagram of chromosomes 3 and 10 breakpoints (Case 1). A. Telomeric region of chromosome 3 from 3p26.1 to 3p26.3, represented by the grey and white bar at the top of the diagram. Relevant STS markers are in black text, and known genes from the RefSeq database are in blue text. The region corresponding to *CNTN4* is pictured as a blue shaded rectangle. Distances noted were derived from the July 2003 freeze of the Human Genome Project and include: (1) the 1.69 Mb-interval from the 3'-most exon of CHL1 to exon 1 of CNTN4; (2) the 0.69-Mb interval from the 3'-most exon of CNTN6 to exon 1 of CNTN4; (3) the 4.5-Mb interval defined by Cargile et al. (2002) as the smallest interstitial deletion identified in a 3p deletion syndrome case; and (4) the 6.7-cM interval identified by Higgins et al. (2002) in a large pedigree affected with nonspecific mental retardation ($Z_{max} = 9.18$ at marker D3S3050). The bar below the genomic interval shows detail of the region, including exons of *CNTN4* (*in red*). Selected BACs are noted as dark horizontal bars beneath. The breakpoint interval is shown as a red rectangle. The position of each BAC probe relative to the breakpoint on chromosome 3 is indicated below the RPCI-11 library identifier. The distance for the entire genomic interval corresponding to *CNTN4* is indicated below the thin black line. **B.** Chromosome 10 in the region from 10q25.3-10q26.12, represented by the grey and white bar at the top of the diagram. Known genes from the RefSeq database are noted in blue. The 0.46-Mb distance between the 3'-most exon of EMX2 and the 5'-most exon of Rab11-FIP2 was determined by the July 2003 freeze of the Human Genome Project. The lower bar shows details of the interval, with the relative position of known genes in the region in blue text. The breakpoint region is highlighted with a red rectangle. No known genes or spliced ESTs were identified on the BAC spanning the chromosome 10 breakpoint. BACs from the region are identified as dark horizontal lines below the genomic region, with their position relative to the breakpoint on chromosome 10 noted below their RPCI identifiers.

BACs RPCI-11 355F22 (Figures 3 and 4) and RPCI-11 136I7 were found to span the chromosome 10 translocation breakpoint. This chromosome segment lies within a gene-poor region in the band 10q26.11 (Figure 4B). No known or predicted transcripts or spliced ESTs were identified on these clones with the use of the UCSC genome browser. Five distinct ESTs without canonical splice sites were mapped to the BAC. Efforts to extend these fragments *in silico* with the use of the BLAST server at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/blast/) were unsuccessful, and four of these five fragments showed no significant homology to other human ESTs or known genes.

Additional FISH studies were undertaken to evaluate the regions telomeric and centromeric to the 3p translocation (Table 1). These resulted in no evidence of a deletion. Subtelomeric FISH with the use of commercially available probes was also performed to confirm the integrity of the terminal regions of the normal and derivative chromosomes 3 and 10 (Figure 3). FISH mapping of BACs corresponding to the two closest transcripts on the telomeric aspect of the breakpoint, *CHL1* (Close Homologue of L1, also known as *CALL* [MIM 607416]) and *CNTN6* (MIM 607220), each showed the expected pattern of two hybridization signals, and direct sequencing of *CNTN4* in the patient demonstrated no mutations in the coding regions of the remaining, intact transcript (data not shown).

Cases 2-4 – autism and a balanced (4;10) translocation

Clinical Assessment

SG (female, age 7 y 3 mo), WG (male, age 7 y 3 mo), and AG (female, age 6 y 1 mo) were evaluated at the Yale Child Study Center (YCSC) under a research protocol approved by the Yale Human Investigations Committee with informed consent. These were the only offspring of a non-consanguineous union, delivered at 34, 34, and 40 weeks, respectively. There was no history of alcohol consumption or medication use during pregnancy, and no known exposure to teratogens. No other family members were reported to have mental retardation or autism, although a maternal cousin was reported to have attentional and social difficulties.

Assessment of adaptive functioning at YCSC using the Vineland Adaptive Behavior Scales (22) showed marked impairments in communication, daily living skills socialization, and motor skills, with percentile rank scores <0.1 for all three children across all domains. Adaptive composite scores (all with percentile rank <0.1) were as follows: SG, 29; WG, 28; AG, 34.

All children were also evaluated for autistic disorder using the Autism Diagnostic Interview-Revised (ADI-R) (20) and Autism Diagnostic Observation Schedule (ADOS) (26). All three children met criteria for autistic disorder by scores on both scales. ADI scores for SG, WG, and AG, respectively, were as follows: communication and language 14, 19, 9 (threshold score for autism is 7), social interaction 24, 23, 16 (threshold score for autism is 10), stereotyped behavior 3, 4, 4 (threshold score for autism is 3). Two different versions of the ADOS were used (21, 26). ADOS scores for SG and WG,

respectively, were as follows: communication domain 10, 8 (threshold score for autism is 4), social domain 14, 14 (threshold score for autism is 7), total 24, 22 (threshold score for autism is 12). ADOS scores for AG were: communication domain 10 (threshold score for autism is 5), social domain 14 (threshold score for autism is 6), total 24 (threshold score for autism is 12).

Cytogenetic analysis

Cytogenetic analysis revealed a balanced translocation [46, XX, t(4;11)(q25;p15)] in both affected female siblings and their unaffected mother (Figure 5). An apparently identical translocation [46, XY, t(4;11)(q25;p15)] was found in the affected male sibling. The unaffected father had a normal karyotype.

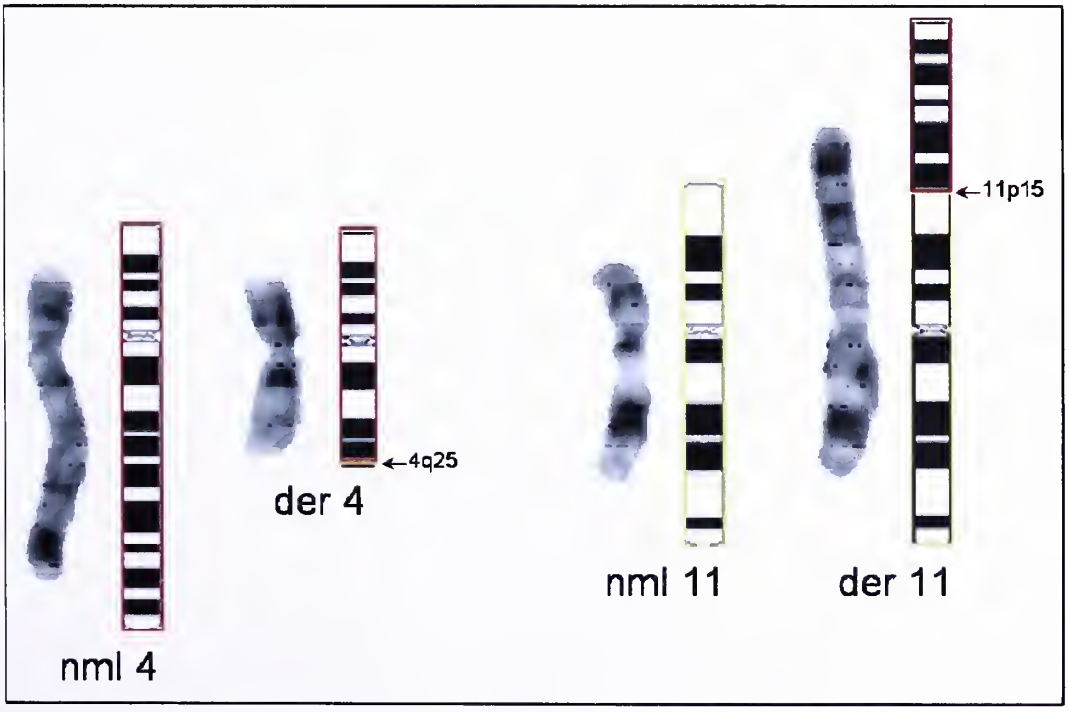

Figure 5. Ideogram of balanced translocation in Cases 2-4: Diagram of chromosomes 4 and 11, showing the balanced translocation t(4;11)(q25;p15). Nml 4 and nml 11 = normal chromosomes 4 and 11. Der 4 and der 11 = derivative chromosomes 4 and 11 with exchange of genetic material. Chromosomes and pieces outlined in red = chromosome 4; chromosomes and pieces outlined in yellow = chromosome 11. Karyotype images of corresponding G-banded chromosomes are shown adjacent to ideograms.

In all four chromosomal rearrangements, the location of the breakpoint on chromosome 4q25 was mapped to within a 148 kb region (Figure 6, Table 2 in Appendix). One known gene (*RPL34*) was localized to this region based on the NCBI Ref Seq database at http://www.ncbi.nlm.nih.gov/RefSeq/. This gene is known to code for a ribosomal protein with moderate expression levels in brain (http://www.ncbi.nlm.nih.gov/UniGene).

The location of the breakpoint on chromosome 11p15.4 was mapped to within a 67 kb region (Figure 6, Table 2). Two predicted genes (*FLJ32752* and *KIAA0409*) map within this interval based on the NCBI Ref Seq database at http://www.ncbi.nlm.nih.gov/RefSeq/. *FLJ32752* encodes a hypothetical protein involved with microtubule motor activity in embryo with no apparent brain expression. *KIAA0409* encodes a brain expressed hypothetical protein (Cerebral protein-1) that appears to belong to a family of uncharacterized eukaryotic proteins which are related to methyltransferases (http://www.sanger.ac.uk/Software/Pfam/iPfam/).

Evaluation of copy number changes using the GeneChip Mapping 100K Set array revealed no areas of deletion or duplication on 4q25 or 11p15.4 (data not shown).

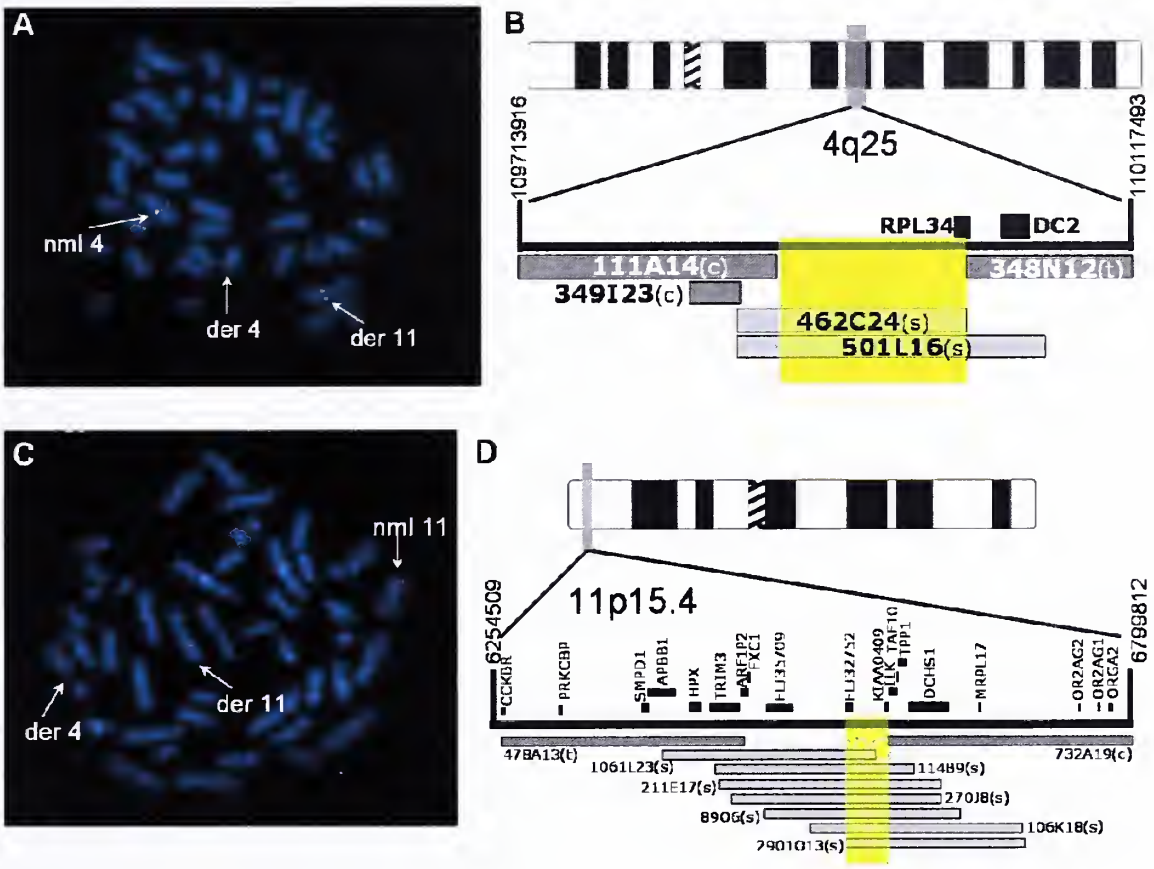

Figure 6. Fine-mapping breakpoints of a (4;11) translocation (Cases 2-4). A. Metaphase cell from mother under fluorescent microscopy with locations of human BAC probes indicated. Aqua (DEAC) probe hybridized to the normal chromosome 4 and the derivative chromosome 11, indicating that this BAC probe is telomeric to the breakpoint on chromosome 4. Red (Cy3) probe hybridized to nl chromosome 4, der chromosome 11, and der chromosome 4, indicating that this BAC probe spans the breakpoint on chromosome 4. **B.** Ideogram of chromosome 4, illustrating the positions of spanning Cy3 (RP11-462C24) and RP11-501L16 probes on a normal chromosome 4. One known gene (RPL34) exists in the 148 kb breakpoint interval (yellow shaded region). C=centromeric, s=spanning, t=telomeric BACs. **C.**. Metaphase cell from mother under fluorescent microscopy with locations of human BAC probes indicated. Aqua (DEAC) and red (Cy3) probes both hybridized to the normal chromosome 11, the derivative chromosome 11, and the derivative chromosome 4, indicating that this BAC probe spans the breakpoint on chromosome 11. Green (FITC) probe was used as a control, hybridizing to 11p15 centromeric to the breakpoint. **D.** Ideogram of chromosome 11, illustrating the positions of spanning (s) DEAC (RP11-106K18), Cy3 (RP11-89O6), and other BAC probes on a normal chromosome 11. Two predicted genes (FLJ32752 and KIAA0409) exist in the 67 kb breakpoint (yellow shaded) region.

Table 2. Cases 2-4: BAC probes investigated using fluorescence *in situ* hybridization[B]

BAC ID	Base position (July 2003 Freeze)	Chromosomal band	Location relative to breakpoint
RP11-297P16	chr4:104,864,485-105,048,993	4q24	centromeric
RP11-41F9	chr4:107,451,468-107,574,511	4q24	centromeric
RP11-713M16	chr4:108,552,432-108,690,056	4q25	centromeric
RP11-286E11	chr4:109,166,532-109,329,848	4q25	centromeric
RP11-343N16	chr4:109,378,308-109,535,867	4q25	centromeric
RP11-558N14	chr4:109,523,271-109,659,770	4q25	centromeric
RP11-349I23	chr4:109,824,239-109,858,719	4q25	centromeric
RP11-111A14	chr4:109,713,916-109,881,949	4q25	centromeric
RP11-501L16	chr4:109,856,719-110,060,415	4q25	spanning
RP11-462C24	chr4:109,856,720-110,008,699	4q25	spanning
RP11-348N12	chr4:110,006,700-110,117,493	4q25	telomeric
RP11-75N20	chr4:110,116,399-110,266,975	4q25	telomeric
RP11-602N24	chr4:111,111,779-111,230,409	4q25	telomeric
RP11-650J17	chr4:114,303,368-114,520,237	4q25	telomeric
RP11-746C23	chr11:12,530,466-12,718,542	11p15.3	centromeric
RP11-483L5	chr11:11,387,947-11,600,728	11p15.4	centromeric
RP11-58H20	chr11:10,534,231-10,699,928	11p15.4	centromeric
RP11-467K18	chr11:8,971,792-9,166,478	11p15.4	centromeric
RP11-321F11	chr11:8,371,551-8,539,322	11p15.4	centromeric
RP11-494M8	chr11:7,764,989-7,959,247	11p15.4	centromeric
RP11-324J3	chr11:7,101,122-7,273,393	11p15.4	centromeric
RP11-413N10	chr11:6,909,050-7,077,179	11p15.4	centromeric
RP11-732A19	chr11:6,588,043-6,799,812	11p15.4	centromeric
RP11-291O13	chr11:6,551,174-6,706,742	11p15.4	spanning
RP11-106K18	chr11:6,520,599-6,704,383	11p15.4	spanning
RP11-89O6	chr11:6,480,882-6,650,346	11p15.4	spanning
RP11-270J8	chr11:6,452,777-6,633,889	11p15.4	spanning
RP11-114B9	chr11:6,438,904-6,611,366	11p15.4	spanning
RP11-1061L23	chr11:6,392,587-6,577,701	11p15.4	spanning
RP11-211E17	chr11:6,441,702-6,633,943	11p15.4	spanning
RP11-478A13	chr11:6,254,509-6,463,471	11p15.4	telomeric
RP11-304C12	chr11:6,247,139-6,438,805	11p15.4	telomeric
RP11-290F24	chr11:6,069,305-6,255,908	11p15.4	telomeric
RP11-451K18	chr11:5,958,461-6,124,142	11p15.4	telomeric
RP11-437G21	chr11:4,204,393-4,407,348	11p15.4	telomeric
RP11-23F23	chr11:4,020,255-4,202,776	11p15.4	telomeric
RP11-438N5	chr11:3,861,913-4,013,518	11p15.4	telomeric
RP11-348A20	chr11:3,678,516-3,843,974	11p15.4	telomeric
RP11-11A9	chr11:3,200,821-3,320,380	11p15.4	telomeric

[B] BAC probes used for FISH analysis: The RPCI-11 identifiers, chromosome number and band, location in base pairs (July 2003 freeze), and position determined relative to the breakpoint(s) are shown for all BACs hybridized. BAC probes noted as centromeric and telomeric yielded two hybridization signals on metaphase spreads. Spanning probes resulted in three hybridization signals.

Mutation screening of *CNTN4*

Based on the finding of a direct disruption of the *CNTN4* gene in a patient (Case 1 above) with an autistic spectrum disorder (PDD-NOS), genomic DNA from 97 AGRE sample patients was screened via dHPLC for mutations in the coding exons of *CNTN4*. These patients all had no known cytogenetic abnormalities and had at least one dysmorphic facial feature in common with patient CT (see Appendix Table A2). The cohort of 97 patients selected included those with a diagnosis of autism (N=78), "broad spectrum" (N=13) and "not quite autism" (NQA, N=6).

Primers were designed to amplify exons and flanking regions of *CNTN4* (see "Amplification of samples" in Materials and Methods section). Forward and reverse primer sequences for each exon, the resulting amplicon size, and PCR annealing temperatures are listed in Appendix Table A1.

Following PCR amplification of each exon, each sample was screened for heterozygous variants using dHPLC methods described above. DHPLC running temperature(s) for each amplicon are listed in Appendix Table A1. Results to-date for this mutation screening of 97 AGRE patients are summarized in Table 3. Based on a comparison of dHPLC chromatogram waveforms for each amplicon, samples with unique-looking variants were identified as candidates for sequencing in order to confirm a suspected heterozygous locus. Based on all samples to-date that have successfully completed dHPLC screening and sequencing of at least one sample representing their unique chromatogram waveform appearance, several heterozygous loci have been confirmed (see Table 3).

Table 3. Mutation screening of *CNTN4* in AGRE cohort (N=97), by amplicon

Amplicon	Coding/ non-coding	# samples completed dHPLC	# samples candidates for sequencing	# samples completed full screen	# heterozygous loci detected
Exon 1	Non-coding	93	13	0	-
Exon 2	Non-coding	84	9	82	0
Exon 3	Coding	96	6	96	0
Exon 4	Coding	46	7	46	0
Exon 5	Coding	81	9	31	0
Exon 6	Coding	97	7	97	1[C]
Exon 7	Coding	79	16	60	0
Exon 8	Coding	91	7	89	1[D]
Exon 9	Coding	90	15	90	0
Exon 10	Coding	75	7	75	0
Exon 11	Coding	90	10	90	0
Exon 12	Coding	95	7	0	-
Exon 13	Coding	96	7	0	-
Exon 14	Coding	97	8	0	-
Exon 15	Coding	93	4	93	0
Exon 16	Coding	80	5	0	-
Exon 17	Coding	97	11	0	-
Exon 18	Coding	97	7	97	0
Exon 19	Coding	97	30	0	-
Exon 20	Coding	65	16	65	0
Exon 21	Coding	97	15	95	37[E]
Exon 22	Coding	85	10	85	0
Exon 23	Coding	97	11	88	0
Exon 24a	Coding	79	19	79	0
Exon 24b	Non-coding	97	13	0	-
Exon 24c	Non-coding	97	12	0	-
Exon 24d	Non-coding	93	7	0	-
Exon 24e	Non-coding	97	7	0	-
Exon 24f	Non-coding	97	16	0	-
Exon 24g	Non-coding	97	19	0	-

[C] Sequencing in both directions confirmed presence of a heterozygous, missense, A→G mutation in base 101 of exon 6 (A723947G). See Figure 7 for details.

[D] Sequencing in one direction indicated a unique variant within intronic sequence following 3' end of exon 8 (IVS8+13 T>G). See Figure 9A.

[E] Sequencing several candidate samples revealed a common intronic SNP before 5' exon 21 (IVS21-8 C>T). All samples with matching dHPLC waveforms were inferred to carry the same variation. See Figure 9B.

One autistic patient was found to have a missense point mutation A723947G (A→G heterozygous change at nucleotide 723947 of *CNTN4*, or nucleotide 101 of exon 6). The dHPLC chromatogram and heterozygous sequence chromatogram are shown in Figure 7. This base change results in an amino acid change (Thr→Ala) at position 128 of the *CNTN4* protein (Figure 7C). This amino acid is located within the second of six immunoglobulin-like C2-type domains of the protein. A pedigree for the Caucasian family of this proband is shown in Figure 8A. Sequencing exon 6 of *CNTN4* in each family member revealed an identical heterozygous A723947G mutation in an autistic sister, unaffected sister, and unaffected mother of the autistic proband (Figure 8B). This exon was also sequenced in 234 Caucasian controls and no mutations were found.

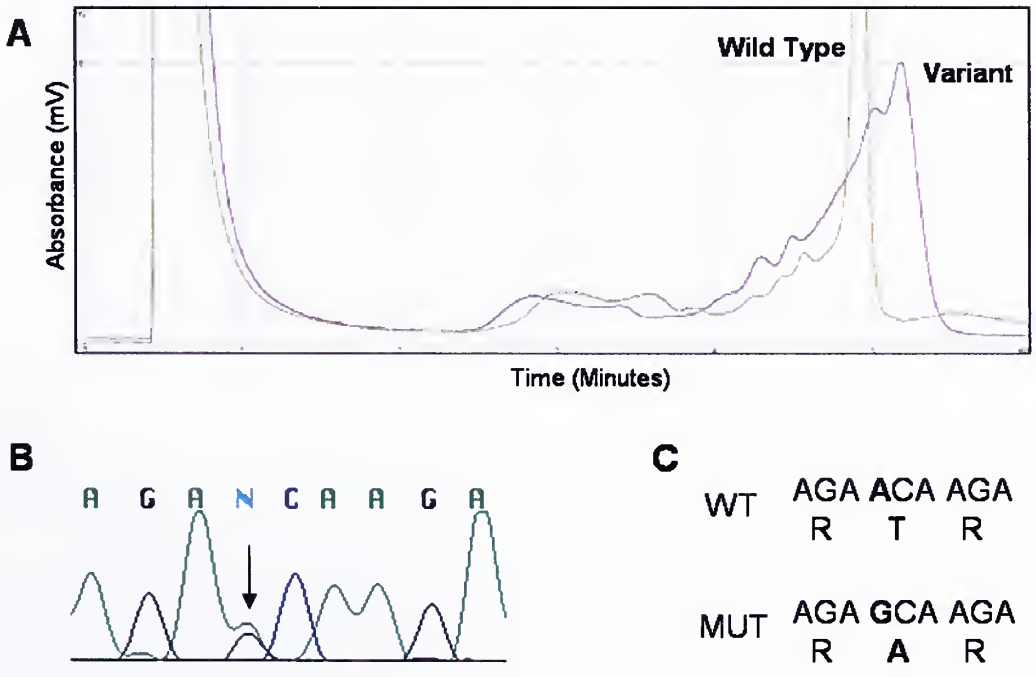

Figure 7. Detection of a missense mutation in exon 6 of *CNTN4* in one autistic patient. A. dHPLC chromatograms show wild type (non-mutated) waveform along with the mutation or variant waveform produced by the DNA of this autistic patient when screening for mutations in exon 6 of *CNTN4*. Based on the waveform variation, this sample was flagged as a candidate for sequencing. **B.** Sequencing chromatogram for this autistic patient, showing the heterozygous nucleotide change in exon 6 (A723947G). **C.** This nucleotide change produces a missense mutation, causing an amino acid change in the *CNTN4* protein (Thr128Ala). This amino acid is known to be in the second of six immunoglobulin-like C2-type domains which make up the protein.

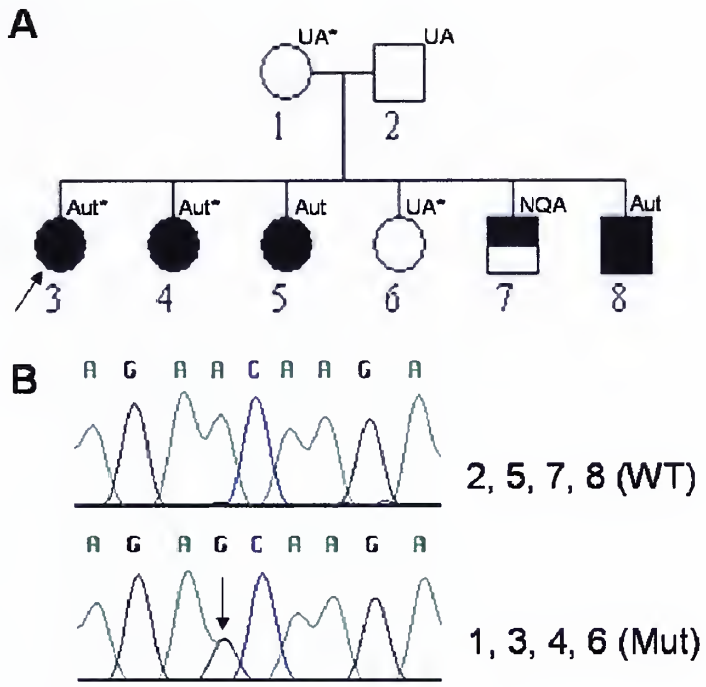

Figure 8. Screening AGRE family for *CNTN4* exon 6 missense mutation. A. Pedigree of AGRE family AU0700 containing index autistic proband (sibling 3 with arrow) who was discovered to have a missense mutation A723947G in exon 6 of *CNTN4* (see Figure 7). Pedigree shows affected status of all members. Aut=autism, NQA=not quite autism, UA=unaffected. Asterisks indicate those individuals who carry a heterozygous A723947G missense mutation: two autistic sisters (3, 4), one unaffected sister (6), and unaffected mother (1). **B.** Sequence chromatograms for family members without the mutation (2, 5, 7, 8) and those with the A→G heterozygous mutation (1, 3, 4, 6).

In addition, a different autistic patient was found to have a unique T→G variant which is located in an intron , 13 nucleotides away from the 3' end of exon 8 (IVS8+13 T>G, Figure 9A). Because this variant is intronic, it is not transcribed into mRNA, and therefore is not involved in coding for an amino acid in the translated protein. A search of online SNP databases (http://www.ncbi.nlm.nih.gov/SNP) does not reveal any recorded information about this particular single nucleotide change, so population frequency is not known.

Finally, 37 autistic, "broad spectrum" and "not quite autism" patients were found to have a C→T variant in an intron, 8 nucleotides away from the 5' end of exon 21

(IVS21-8 C>T, Figure 9B). This non-coding variant is recorded in online SNP databases (refSNP ID 6800354) as having an average allele frequency of 0.17.

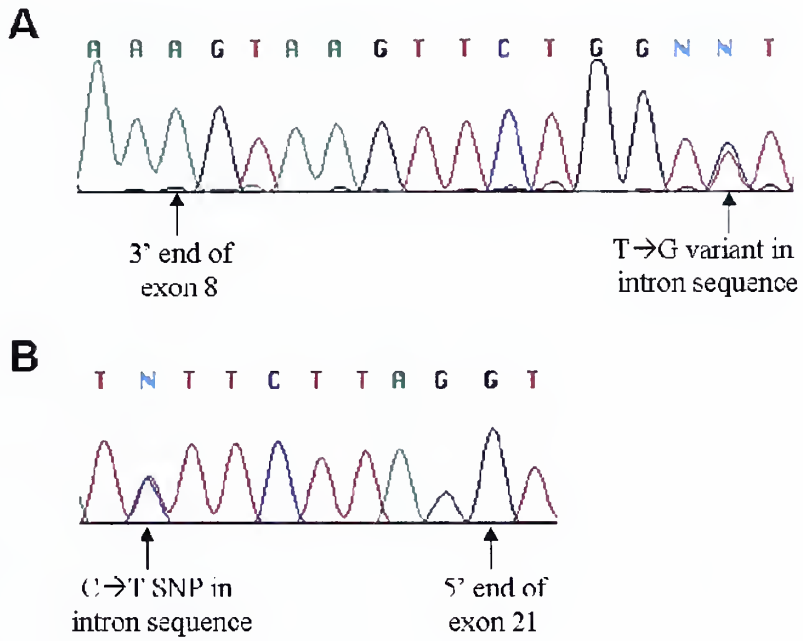

Figure 9. Heterozygous intronic variants in *CNTN4* detected by dHPLC and confirmed with sequencing. A. Sequencing chromatogram of one autistic patient with heterozygous intronic variant, located 13 nucleotides away from the 3' end of *CNTN4* exon 8 (IVS8+13 T>G). This variant has not been previously reported in online SNP databases. **B.** Sequencing chromatogram from one of 37 autistic, broad spectrum, and NQA patients with identical dHPLC chromatograms (not shown). The sequence reveals a heterozygous intronic SNP, located 8 nucleotides away from the 5' end of *CNTN4* exon 21 (IVS21-8 C>T). This SNP has been previously reported in the general population (refSNP ID 6800354), with an estimated (T) allele frequency of 0.17.

DISCUSSION

Over the last decade, many studies have aimed to identify genetic susceptibility loci for developmental neuropsychiatric disorders. The importance of molecular genetic findings lie in their ability to increase our understanding of the underlying pathology, improve diagnosis and genetic counseling, and eventually lead to new strategies of prevention and treatment (9). While linkage studies have been very successful for

mapping disease loci corresponding to rare disorders with Mendelian inheritance patterns, the field of psychiatric genetics has been less successful in attempts to find relevant genes for relatively common, complex disorders using this method. For autism and ASDs, the several independent linkage studies published to date have produced nominal to suggestive linkage evidence on varied chromosomes, with few regions of overlap between studies. Genetic and phenotypic heterogeneity in ASDs, combined with potentially inadequate sample sizes, has limited the power of non-parametric linkage methods to detect genes of small effect in these disorders.

An alternative approach to identifying genetic susceptibility loci for autism involves the characterization of rare affected individuals who present with chromosomal abnormalities. Candidate genes at or near chromosomal breakpoints may be further investigated by mutation analysis and association studies. This approach has proven successful in identifying mutations in *NLGN4* as rare genetic variants conferring susceptibility to ASDs (14, 15). Important genetic variants such as those in *NLGN4* would not be detectable using other more traditional methods of disease gene identification (e.g.. linkage), owing to the insufficient power of these methods to detect variants that occur relatively infrequently. Discovering such variants, even if rare, is enormously important in progressing toward a thorough understanding of genetic risk for complex, multigenic disorders such as many of those encountered in psychiatry. Furthermore, defining the role of genetic factors will allow us to better evaluate the role of non-genetic (i.e. environmental) factors that certainly contribute to disease susceptibility.

Case 1

Based on standardized testing at the YCSC, CT met cutoff scores for a DSM-IV diagnosis of Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS), which is categorized as one of several autism spectrum disorders (ASDs) (25). In addition, CT's clinical presentation, including developmental delay, growth retardation, and ptosis, in combination with the known cytogenetic abnormality, suggested a diagnosis of 3p deletion syndrome. Appendix Table A2 lists the features of this syndrome, indicating those present in CT. Those physical signs not present in this case (e.g. trigonocephaly, microcephaly, micrognathia) are not ubiquitous findings in other reported cases in the literature. The patient's particular cognitive findings are mirrored in at least one other case in which the patient was found to have a full scale IQ in the borderline mental retardation (MR) range with what appeared to be significant weaknesses in adaptive functioning (36). In addition, a marked delay in early language acquisition, as seen in the present case, frequently has been noted anecdotally. However, a more detailed comparison of CT's cognitive findings with other reported cases is not possible, as a result of the informal descriptions of developmental delay that characterize the 3p deletion syndrome literature. From reports in the literature, it is apparent that phenotypic variability in all aspects of the syndrome appears to be the rule rather than the exception (37-43) .

CT meets criteria for borderline MR, PDD-NOS, and 3p deletion syndrome. In this case, the finding of disruption of *CNTN4* is of particular interest since it maps both within the minimal candidate region for 3p deletion syndrome defined by Cargile et

al.(44) and within a 6.7-cM region found linked to nonspecific mental retardation in large kindred (Zmax = 9.18 at marker D3S3050) (45) (Figure 4A). This gene encodes a neuronal adhesion molecule that is a member of the immunoglobulin (Ig) superfamily of genes. Similar to other members of the contactin family, which includes F3/F11/contactin (MIM 600316), *TAG-1* (*CNTN2* [MIM 190197]), *BIG-1* (*CNTN3* [MIM 601325]), *NB-2* (*CNTN5* [MIM 607219]), and *NB-3* (*CNTN6*), full-length *CNTN4* consists of six immunoglobulin (Ig) domains and four fibronectin III (FNIII) domains and is anchored to the cell membrane via a glycosylphosphatidyl-inositol (GPI) domain. This family of proteins has been shown to be involved in axon growth, guidance, and fasciculation (46-49) and have been proposed to have a role in synaptic plasticity (50). While a specific role for *CNTN4* in neuronal development has not yet been clarified, knockouts of homologous neuronal adhesion molecules in mouse have resulted in viable mutants demonstrating morphological, neurological, and behavioral abnormalities (51-54).

While three isoforms of *CNTN4* are described in the NCBI RefSeq database (http://www.ncbi.nlm.nih.gov/RefSeq), the identified breakpoint in this case disrupts only the longest variant, which is highly expressed in human brain, particularly in cerebellum, thalamus, amygdala, and cerebral cortex. The intermediate-sized isoform has not been found to be expressed in the CNS, while the third, shortest variant, lacking all six Ig and three FNIII domains, shows low levels of brain expression on northern blot and RT-PCR analysis (49, 55). Given the unavailability of brain tissue, it was not possible to directly test whether the current translocation altered expression of this short brain-expressed variant, despite the transcript remaining structurally intact. However, the likelihood of this short variant being able to compensate for the disruption of the full-length isoform is

unlikely for several reasons. First, this short variant lacks all Ig and several FNIII domains which, in the full-length isoform, participate in a complex pattern of homophilic and hetetrophilic interactions that are required for axon growth and pathfinding. Second, the short isoform displays low levels of brain expression that are confined to the cerebral cortex, with little expression overlap with that of the longest isoform. Finally, the phenotypic features seen with CT overlap with those seen in cases where the terminal region of chromosome 3 is deleted. For these reasons, it is most likely that interruption of the longest *CNTN4* isoform alone is sufficient to result in functional haploinsufficiency at this locus.

The region telomeric to *CNTN4* contains two genes (*CNTN6* and *CHL1*) that also encode neuronal cell adhesion molecules with Ig and FNIII domains (Figure 4A). Given their similarity and proximity to the breakpoint locus in this case, and noting a report of a patient with nonspecific MR and a balanced translocation disrupting *CHL1* (56) as well as reports of *CHL1* knockout mice displaying abnormalities of behavior and axonal organization, it seems possible that the chromosomal translocation disrupting *CNTN4* in this case may functionally disrupt these neighboring genes and lead to CT's phenotype. While such position effects are certainly possible (57), one should note that in the single reported case of MR associated with *CHL1* disruption, the patient did not display any dysmorphic features associated with the 3p deletion syndrome (56). In addition, the 3p-minimal candidate region suggested by Cargile et al.(44) includes *CNTN4*, but neither *CNTN6* nor *CHL1* (Figure 4A).

Finally, it is noteworthy that a previous analysis of an extended pedigree identified evidence for a recessive locus conferring nonspecific MR within a 6.7 cM

linkage interval, and this region contained *CNTN4* (Figure 4A) (45). The finding of direct structural disruption of *CNTN4* in CT together with these additional lines of evidence all converge upon the most likely scenario and main conclusion from this case: disruption of a single copy of *CNTN4* is sufficient to cause key aspects of the 3p deletion syndrome phenotype, including developmental delay.

Mutation screening of CNTN4

Given the finding and evidence described above, the disruption of *CNTN4* in a patient with borderline MR, an autism spectrum phenotype (PDD), and 3p deletion syndrome provides a rationale for mutation screening of *CNTN4* in selected patients with developmental delay and dysmorphic features who have normal karyotypes. DNA from a cohort of 97 patients with autism and ASDs from the AGRE repository was screened for mutations in the exons of *CNTN4*. Each of these patients had one or more dysmorphic physical features in common with the patient described in Case 1 (see Appendix Table A2 and Figure 2) and no known cytogenetic abnormalities. One Caucasian autistic patient from this cohort was found to have a missense (amino acid changing) heterozygous point mutation in exon 6 of *CNTN4* (Table 3 and Figure 7), which was not found in any of 234 Caucasian normal controls who were sequenced for exon 6. In this case, it is conceivable that the mutation T723947G, resulting in a Thr128Ala amino acid substitution, may alter the function of the *CNTN4* protein, as it alters a highly conserved amino acid in the second of 6 functional Ig-like C2-type domains.

The proband in whom this mutation was detected was one of six siblings in a multiplex ASD family (Figure 8A). Sequencing each member of the immediate family for the T723947G mutation in *CNTN4* revealed three additional family members with the mutation (one autistic sister, one unaffected sister, and the unaffected mother). It appears that this mutation does not completely co-segregate with an autism or ASD diagnosis in this family. While such co-segregation would lend convincing evidence to the argument that this particular mutation is associated with autism in this family, lack of co-segregation does not necessarily exclude this possibility. One probable explanation is that this missense mutation may increase susceptibility to autism or ASDs, but is not the sole cause. The genetic change resulting from this mutation may be compounded by other genetic and non-genetic factors in those affected, such as a secondary mutation, or environmental factors, such as exposure to pathogens or teratogens, that increase the likelihood of developing autism in this family. Second, the genetic change may be one of major effect but may have reduced penetrance, explaining why there is no manifestation of autism in the mother or one of the sisters with the mutation. Incomplete penetrance appears to be the rule in psychiatric diseases, as the concordance rates for autism in monozygotic twins (who share a similar environment and all of their genetic material) varies from 40-60% and from 0-25% in dizygotic twins (who share a similar environment and half of their genes, on average) (1, 2, 3, 16). Finally, the phenotypic manifestations of the identified mutation in this family may be heterogeneous. In the example of the recently discovered association between idiopathic autism and mutations in *NLGN4* (14) (see Introduction), it was later found that mutations in this gene also manifested as X-linked mental retardation with or without autism or PDD in a large family, suggesting

that some types of autistic disorder and mental retardation may have common genetic origins (15). In a similar manner, a closer phenotypic evaluation of this family may suggest that the missense mutation discovered here may co-segregate with some other aspect of their presentations.

Keeping with the assumption that this T723947G mutation in *CNTN4* increases susceptibility to ASDs in this family, it is more difficult to explain why there are three affected siblings (two with autism and one "NQA") that do not carry this mutation. It seems more likely that other, more relevant (genetic or other) risk factors for ASD are shared in this family, and that the particular *CNTN4* mutation detected here is of very little, if any, significance toward autism susceptibility.

In addition to this genetic change within an exon of *CNTN4*, two different intronic genetic variations were observed in different patients. Although this study did not set out to examine intronic, or non-coding, sequence of *CNTN4*, the purposeful design of PCR primers to include sequence flanking the intron/exon boundaries, or splice sites, invariably included several base pairs of intron sequence. In one autistic patient, a heterozygous intronic variant (IVS8+13 T>G) was found (Table 3, Figure 9A). There is no record of this variation being reported in any online SNP databases, either because it is truly rare or because the SNP has yet to be recorded. Due to its distance from the exon 8 splice site, it is unlikely that this mutation would interfere with normal transcription. However, it may be revealing to check for this unreported intronic variant in control subjects and members of the proband's family.

The second heterozygous intronic variant (IVS20-8 C>T) was found in 37 of 95 AGRE patients (Table 3, Figure 9B). According to the online SNP databases, this variant

is common in the general population (refSNP 6800354), with an estimated (T) allele frequency of 0.17. Due to its relatively high prevalence and location in non-coding sequence, it is unlikely that this SNP contributes to autism susceptibility in any significant way. Such variations are relatively common in the human genome. Current estimates are that SNPs (by definition, having a minor allele frequency of at least 1% in the general population) are found at the rate of 1 in every 200–300 bp in the genome (58, 59). While most SNPs in the human genome are found outside of coding sequence, SNPs found within a coding sequence are of particular interest because they are more likely to alter the biological function of a protein.

Based on these initial mutation screening results for *CNTN4*, it would be valuable to complete mutation screening for all translated exons for the entire initial cohort of AGRE patients. With recent technological advances, screening for mutations has become faster and cheaper since the start of this study. As a result, it may be feasible to expand this initial cohort to include additional AGRE patients as well as patients with other phenotypic similarities to our index case (Case 1), namely patients with MR and/or non-verbal learning disability with a delay in language acquisition.

Cases 2-4

One male and two female siblings meeting Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV), as well as Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS), criteria for autistic disorder and their unaffected mother were found to have a balanced translocation

t(4;11)(q25;p15). One breakpoint was mapped to within a 67 kb region on chromosome 11p15.4 (Figure 6B). The location of this breakpoint is approximately 13 Mb telomeric to a recently reported area of interest (multipoint LOD =1.72) in autism spectrum disorders based on a genome-wide screen of 345 families (10). The second breakpoint was mapped to within a 148 kb region on chromosome 4q25 (Figure 6D), mapping a considerable distance (28 Mb) from a region of suggestive linkage (multipoint LOD=2.24) in the same genome-wide scan.

While several case reports of abnormalities involving chromosome 4q have been noted in the literature (60-62), none of these breakpoints overlap with or even fall within 20 Mb of the 4q breakpoint reported here. Several association studies of candidate genes near our breakpoint on 11p15.4 may complement our evidence for a possible susceptibility gene at or near this locus. Herault et al. report the identification of a positive association between autism and a particular genotype at two different c-Harvey-ras (HRAS) oncogene markers located on 11p15.5 (63-65). This association at one such marker was later replicated in a second population of autistic subjects (66). The product of HRAS is involved in the proliferation and differentiation of neural cells, and therefore could play a role in the etiology of autism (65). In addition, it is interesting to note that the Ras protein is an upstream regulator of the ERK MAPK signaling cascade, which is widely involved in CNS synaptic plasticity (67, 68). With this crucial role of Ras in one of the most studied pathways in learning and memory, a mutation affecting this protein could adversely affect important synaptic changes that occur during normal development and result in developmental delay. In fact, mutations of several genes coding for proteins that interact with Ras (NF1) and other proteins in the ERK pathway (RSK2, CPB) have

been shown to cause several developmental disorders (mental retardation in neurofibromatosis, Coffin-Lowry Syndrome MR, and Rubinstein-Taybi Syndrome MR, respectively) (68-70). Yamagata et al. examined the human secretin locus (SCT), adjacent to HRAS at 11p15.5, and found three out of 75 Caucasian and Japanese autistic patients who displayed unique, non-conservative coding mutations that were not present in controls (71). The breakpoint on 11p in our family is located approximately 5.9 Mb from these previously reported candidate genes.

While our breakpoints fall outside of previously reported linkage peaks and associated candidate gene intervals, several cases have been described in which chromosomal abnormalities separated by large genomic distances from disease-related genes may still disrupt gene expression by epigenetic alterations and lead to the disease phenotype (57, 72, 73). Mechanisms by which this may occur include the dissociation of a regulatory element from a disease gene or the transposition of a gene from its usual chromatin environment which may result in inadequate or mistimed expression of a functional protein. In this manner, it is still possible that one or both of the breakpoints identified in this multiplex family may correlate with increased susceptibility to autism, even if they do not structurally disrupt a transcript. As illustrated by the few cases of X chromosome deletions that led to the discovery of a correlation between mutations in *NLGN4* and autism (13, 14), studying chromosomal abnormalities, even when breakpoints are outside of known linkage or candidate gene intervals, can allow us to detect low frequency disease mutations that may not be detected via other methods.

There is still the possibility that one of the breakpoints observed in this family physically disrupts a transcript within the identified breakpoint regions or disrupts a

regulatory region for a nearby transcript. One predicted gene identified in the interval on chromosome 11 (*KIAA0409*) is of particular interest for possible relevance toward the affected status of the three children in this family. This predicted gene is noted to encode a hypothetical protein (Cerebral protein-1) that appears to belong to a family of uncharacterized eukaryotic proteins which are related to methyltransferases. It is interesting to note that abnormalities in repressing the transcription of methylated DNA in brain have been shown to lead to an autism-spectrum phenotype in females. Specifically, mutations in the gene encoding MeCP2 have been shown to cause Rett's syndrome (74, 75). MeCP2 is a protein that has been shown to be particularly abundant in brain, binding specifically to methylated DNA to repress transcription. Since the *KIAA0409* protein has a putative function related to DNA methylation in brain, it is conceivable that a functional alteration of this gene may confer an autism phenotype by mechanisms similar to that described for MeCP2. Further investigation into whether this gene is functionally disrupted in this family and possibly mutation screening of this gene in ASD patients is warranted.

The putative association between the currently reported translocation and autism in this family should not be dismissed merely because the mother does not show features of autism. One possible explanation is that the translocation is not solely responsible for the siblings' autism. The genetic change resulting from the rearrangement may be compounded by other genetic and non-genetic factors, such as a secondary mutation, or environmental factors, such as exposure to pathogens or teratogens, that increase the likelihood of developing autism. Second, the abnormality may have reduced penetrance, explaining why there is no manifestation of autism in the mother. Finally, it is not

possible to exclude epigenetic factors as an explanation for the unaffected mother. The normal maternal imprinting of a genetic locus on chromosome 4 or 11 may be affected by the balanced chromosomal translocation such that maternal transmission of this locus may have dramatically increased susceptibility to autism in her offspring. A precedent for such a potential mechanism in has previously been implicated in psychiatry; maternal transmission of interstitial duplications of 15q11-13 is the most frequent chromosomal abnormality in ASD (76, 77);

In summary, the presence of three affected children with classic autism all carrying an identical rearrangement raises the prospect of a causal relationship in this family. This finding suggests that further annotation of these genomic intervals and mutation screening of prioritized candidate genes at these breakpoints should be undertaken.

Conclusions

This study has identified a gene, *Contactin 4*, whose haploinsufficieny appears sufficient to confer major aspects of the 3p deletion syndrome phenotype, including developmental delay. Initial mutation screening of this gene in a cohort of patients with autism spectrum disorders suggests that *CNTN4* may have relevance in the pathogenesis of rare cases of autism spectrum disorders. In addition, a brain-expressed gene was identified near the translocation breakpoint in a multiplex autistic family that appears to be a good candidate for mutation screening in autism.

As demonstrated by the cases of chromosomal rearrangements analyzed in this study, the investigation of genotypic outliers in complex disease such as developmental

neuropsychiatric disorders is a valuable and worthwhile approach toward elucidating genetic risk factors. This approach has shown success in identifying rare genetic variants conferring susceptibility to ASDs (*NLGN4)*, a task in which more traditional approaches such as whole genome linkage analysis would fail. Despite being rare, these susceptibility variants are invaluable because they may provide insight into the underlying biology of the disorders and may provide unexpected clues about where to look to discover more common risk factors.

APPENDIX

Table A1. Conditions for PCR amplification and dHPLC mutation screening of *CNTN4* exons

Exon	Forward Primer (5'→ 3')	Reverse Primer (5'→ 3')	Amplicon size (bp)	Annealing Temp (°C)	DHPLC Temp (°C)
1	TGCCTATGAGCAGGTCCATAC	CGACCCACTGCAGATATTCC	234	56.0	59.0
2	ACAGGTAAGCATCACCTACCC	TCTGTAAGGGAAGGTTTTCAATG	233	58.1	55.0
3	TTTGAGTAAAGATAAACTTGCAGAGAC	AAAGTGAAGGCCCACTTTTG	300	56.8	52.0, 57.0
4	TGGAAGGGATTCTTCATTTTG	CAGAAACTATATGCATGGAAACAC	215	58.1	55.0
5	GATGTCACCATTGCTGGGTAG	AAAATCTAGTCTCTAAAATGAAACTGC	439	59.1	53.0, 58.0
6	TTTGAAATAGTGGTACTCTCCCTTG	GGAGGGAGAACATTGAGGAAG	243	58.1	54.0, 59.0
7	TGGGCTGAAAAAGAGTCATTG	TACCTCCACATTCATCCTGTG	382	59.1	57.0
8	CATTCTGCATTAACAGAAAAATGAG	GAAGGGAGGATGAGAAGGAA	225	58.1	53.0, 58.0
9	ACCAAGAAGGAAGCGTTTAGG	CCTATTCTGCACCATCTCTCC	375	56.8	57.0
10	TGAGTACACACTGAATATACACCTTTC	GGCAATGAGGCTATATTAAACG	250	58.1	56.0
11	TGGTAAAATTGCCTTAAAGTCTCAG	TGGATTGAACATTTATGCTTGG	288	58.1	56.0
12	ATCTCATGGAAGGCTGTCTTG	AATTCCCAAATCCAGAAGGAC	392	58.1	56.0
13	TGAGATAAATTTCACCATTTTTGG	TCGGTAGATTACATTATTAGACACGAG	241	58.1	56.0
14	TTTCCACACAAGCTCACAGAC	TTTATTGTGCAAAAAGGAGCAG	382	58.1	53.0, 58.0
15	CCATTAGTCACCGTGTCCTTC	GTTTCTGGAACTCTGCGCTAC	280	58.1	58.0
16	TTGAATACACTGCTCCAATTCTG	GCTGAGTTCCCAATGTCTCTG	245	56.0	56.0, 61.0
17	TTCCTAACGTAATCTCTGCATTTG	CAAGAATTCAAACGCTTTTCC	294	60.0	60.0
18	TTGCCGCCTGACATAACTTG	CCCAGCTATGTCTAACATGGAG	191	60.0	55.0, 60.0
19	CGATTTTTGCATTTGAGTGAAAC	TGTACCATGCGAAATTGATTG	389	58.1	55.0, 60.0
20	AATGCTAGTATCTACAATCCTGTCC	CAAGGGGCATAGACTCTCTTTC	233	58.1	53.0, 58.0
21	CCTTTGTCCTGATAGCTCTGC	TTGGGATATGTGTGATGAATCC	380	58.1	58.0
22	GGAGATATTCCTCAGAATCCATTG	CACTCACAGCACGATTAGGTG	227	58.1	56.0
23	TCCATATTTGGGACCTTTGTG	TAAGACAGGCAGGAGAAAAGG	270	58.1	59.0
24a	TCATTTCCAGTAGAATCCATGC	TGCATATAAGGACATATGTAAACTGC	480	58.1	55.0, 60.0
24b	TGCTTCTTTAGGAATGGCATTATAC	TTTGCATACAACAGCCTTGAG	398	58.1	50.0, 55.0
24c	TGTAAAATGCTATGGGCAACAC	AACTTTGGGTCTATGTTACTACAGG	500	58.1	53.0, 58.0
24d	TTGAGATGGTTTCAGGAATGG	GATTCAAAAGAAGGGCTCTATTATTAC	386	58.1	51.0, 56.0
24e	AAGTAATTCTCTGTCAATGGATTCG	TTGCCTTTTATGTCACTGTCC	383	58.1	53.0
24f	TTCTGTTTTGCAATAGGTAGCC	GGAACACAGCTTGAAATACTGAC	397	58.1	54.0
24g	TAAACATGCTGCCACCTTTTC	GGTAAGCTAACAAATTCCCTGAC	458	58.1	54.0

Table A2. Selected features of 3p deletion syndrome

Features of 3p- syndrome	Percent Affected[A]	Present (+) or Absent (-) in Case 1
Developmental Delay	86%	+
Post natal growth retardation	86%	+
Ptosis	77%	+
Low birth weight	73%	-
Malformed ears	68%	+
Long Philtrum	68%	-
Broad Nasal Bridge	64%	+
Hypotonia	50%	+
Microcephaly	50%	-
Micrognathia	50%	-
Epicanthal Folds	41%	+
Post Axial Polydactyly	40%	-
Hyperteliorism	36%	+
Feeding problems	29%	+
Clinodactyly	27%	+
Trigonocephaly	23%	-
Downturned corners of mouth	24%	+
Synophrys	18%	+
Low Frontal Hairline	14%	+

[A] Frequency data is derived from the most recent and largest case review in the literature (43) consisting of 22 total patients

REFERENCES

1. Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., and Rutter, M. 1995. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.* **25**:63-77.

2. Folstein, S., and Rutter, M. 1977. Infantile autism: a genetic study of 21 twin pairs. *J. Child Psychol. Psychiatry* **18**:297-321.

3. Lauritsen, M., and Ewald, H. 2001. The genetics of autism. *Acta Psychiatr. Scand.* **103**:411-427.

4. Bailey, A., Palferman, S., Heavey, L., and Le Couteur, A. 1998. Autism: the phenotype in relatives. *J. Autism Dev. Disord.* **28**:369-392.

5. Szatmari, P., Jones, M.B., Zwaigenbaum, L., and MacLean, J.E. 1998. Genetics of autism: overview and new directions. *J. Autism Dev. Disord.* **28**:351-368.

6. Rutter, M. 2000. Genetic studies of autism: from the 1970s into the millennium. *J. Abnorm. Child Psychol.* **28**:3-14.

7. Yonan, A.L., Palmer, A.A., Smith, K.C., Feldman, I., Lee, H.K., Yonan, J.M., Fischer, S.G., Pavlidis, P., and Gilliam, T.C. 2003. Bioinformatic analysis of autism positional candidate genes using biological databases and computational gene network prediction. *Genes Brain Behav.* **2**:303-320.

8. Spielman, R.S., and Ewens, W.J. 1996. The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* **59**:983-989.

9. Maestrini, E., Paul, A., Monaco, A.P., and Bailey, A. 2000. Identifying autism susceptibility genes. *Neuron* **28**:19-24.

10. Yonan, A.L., Alarcon, M., Cheng, R., Magnusson, P.K., Spence, S.J., Palmer, A.A., Grunn, A., Juo, S.H., Terwilliger, J.D., and Liu, J. et al. 2003. A genomewide screen of 345 families for autism-susceptibility loci. *Am. J. Hum. Genet.* **73**:886-897.

11. Collins, F.S. 1992. Positional cloning: let's not call it reverse anymore. *Nat. Genet.* **1**:3-6.

12. Ropers, H.H., and Hamel, B.C. 2005. X-linked mental retardation. *Nat. Rev. Genet.* **6**:46-57.

13. Thomas, N.S., Sharp, A.J., Browne, C.E., Skuse, D., Hardie, C., and Dennis, N.R. 1999. Xp deletions associated with autism in three females. *Hum. Genet.* **104**:43-48.

14. Jamain, S., Quach, H., Betancur, C., Rastam, M., Colineaux, C., Gillberg, I.C., Soderstrom, H., Giros, B., Leboyer, M., and Gillberg, C. et al. 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat. Genet.* **34**:27-29.

15. Laumonnier, F., Bonnet-Brilhault, F., Gomot, M., Blanc, R., David, A., Moizard, M.P., Raynaud, M., Ronce, N., Lemonnier, E., and Calvas, P. et al. 2004. X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am. J. Hum. Genet.* **74**:552-557.

16. Yan, J., Oliveira, G., Coutinho, A., Yang, C., Feng, J., Katz, C., Sram, J., Bockholt, A., Jones, I.R., and Craddock, N. et al. 2004. Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. *Mol. Psychiatry.*

17. Chih, B., Afridi, S.K., Clark, L., and Scheiffele, P. 2004. Disorder-associated mutations lead to functional inactivation of neuroligins. *Hum. Mol. Genet.* **13**:1471-1477.

18. Veenstra-VanderWeele, J., and Cook, E.H. 2004. Molecular genetics of autism spectrum disorder. *Mol. Psychiatry.*

19. Lichter, P., Tang, C.J., Call, K., Hermanson, G., Evans, G.A., Housman, D., and Ward, D.C. 1990. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* **247**:64-69.

20. Lord, C., Rutter, M., and Le Couteur, A. 1994. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* **24**:659-685.

21. Lord, C., Risi, S., Lambrecht, L., Cook, E.H.,Jr, Leventhal, B.L., DiLavore, P.C., Pickles, A., and Rutter, M. 2000. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30**:205-223.

22. Sparrow, S.S. 1984. *Vineland adaptive behavior scales : interview edition, survey form manual.* American Guidance Service. Circle Pines, Minn. 301pp.

23. Geschwind, D.H., Sowinski, J., Lord, C., Iversen, P., Shestack, J., Jones, P., Ducat, L., Spence, S.J., and AGRE Steering Committee. 2001. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.* **69**:463-466.

24. Anonymous 1992. *International statistical classification of diseases and related health problems.* 10th revision edition. World Health Organization. Geneva.

25. Anonymous 1994. *Diagnostic and statistical manual of mental disorders : DSM-IV.* 4th edition. American Psychiatric Association. Washington, DC. 886pp.

26. Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., and Schopler, E. 1989. Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J. Autism Dev. Disord.* **19**:185-212.

27. DiLavore, P.C., Lord, C., and Rutter, M. 1995. The pre-linguistic autism diagnostic observation schedule. *J. Autism Dev. Disord.* **25**:355-379.

28. Oefner, P.J., Huber, C.G., Umlauft, F., Berti, G.N., Stimpfl, E., and Bonn, G.K. 1994. High-resolution liquid chromatography of fluorescent dye-labeled nucleic acids. *Anal. Biochem.* **223**:39-46.

29. Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L., and Oefner, P.J. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**:996-1005.

30. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., and Kalyanaraman, N. et al. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22**:231-238.

31. Spiegelman, J.I., Mindrinos, M.N., and Oefner, P.J. 2000. High-accuracy DNA sequence variation screening by DHPLC. *BioTechniques* **29**:1084-90, 1092.

32. Choy, Y.S., Dabora, S.L., Hall, F., Ramesh, V., Niida, Y., Franz, D., Kasprzyk-Obara, J., Reeve, M.P., and Kwiatkowski, D.J. 1999. Superiority of denaturing high performance liquid chromatography over single-stranded conformation and conformation-sensitive gel electrophoresis for mutation detection in TSC2. *Ann. Hum. Genet.* **63 (Pt 5)**:383-391.

33. Tibanyenda, N., De Bruin, S.H., Haasnoot, C.A., van der Marel, G.A., van Boom, J.H., and Hilbers, C.W. 1984. The effect of single base-pair mismatches on the duplex stability of d(T-A-T-T-A-A-T-A-T-C-A-A-G-T-T-G) . d(C-A-A-C-T-T-G-A-T-A-T-T-A-A-T-A). *Eur. J. Biochem.* **139**:19-27.

34. Ke, S.H., and Wartell, R.M. 1993. Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA; determination by temperature-gradient gel electrophoresis. *Nucleic Acids Res.* **21**:5137-5143.

35. O'Donovan, M.C., Oefner, P.J., Roberts, S.C., Austin, J., Hoogendoorn, B., Guy, C., Speight, G., Upadhyaya, M., Sommer, S.S., and McGuffin, P. 1998. Blind analysis

of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics* **52**:44-49.

36. Angeloni, D., Lindor, N.M., Pack, S., Latif, F., Wei, M.H., and Lerman, M.I. 1999. CALL gene is haploinsufficient in a 3p- syndrome patient. *Am. J. Med. Genet.* **86**:482-485.

37. Brand, A., Reifen, R.M., Armon, Y., Kerem, E., Horenstein, E., and Gale, R. 1987. Double mitral valve, complete atrioventricular canal, and tricuspid atresia in chromosomal 3P-syndrome. *Pediatr. Cardiol.* **8**:55-56.

38. Phipps, M.E., Latif, F., Prowse, A., Payne, S.J., Dietz-Band, J., Leversha, M., Affara, N.A., Moore, A.T., Tolmie, J., and Schinzel, A. 1994. Molecular genetic analysis of the 3p- syndrome. *Hum. Mol. Genet.* **3**:903-908.

39. Drumheller, T., McGillivray, B.C., Behrner, D., MacLeod, P., McFadden, D.E., Roberson, J., Venditti, C., Chorney, K., Chorney, M., and Smith, D.I. 1996. Precise localisation of 3p25 breakpoints in four patients with the 3p-syndrome. *J. Med. Genet.* **33**:842-847.

40. Green, E.K., Priestley, M.D., Waters, J., Maliszewska, C., Latif, F., and Maher, E.R. 2000. Detailed mapping of a congenital heart disease gene in chromosome 3p25. *J. Med. Genet.* **37**:581-587.

41. Tazelaar, J., Roberson, J., Van Dyke, D.L., Babu, V.R., and Weiss, L. 1991. Mother and son with deletion of 3p25-pter. *Am. J. Med. Genet.* **39**:130-132.

42. Knight, L.A., Yong, M.H., Tan, M., and Ng, I.S. 1995. Del(3) (p25.3) without phenotypic effect. *J. Med. Genet.* **32**:994-995.

43. Benini, D., Vino, L., Vecchini, S., and Fanos, V. 1999. 46, XY, del (3) (pter-->p25) syndrome: further delineation of the clinical phenotype. *Eur. J. Pediatr.* **158**:955-957.

44. Cargile, C.B., Goh, D.L., Goodman, B.K., Chen, X.N., Korenberg, J.R., Semenza, G.L., and Thomas, G.H. 2002. Molecular cytogenetic characterization of a subtle interstitial del(3)(p25.3p26.2) in a patient with deletion 3p syndrome. *Am. J. Med. Genet.* **109**:133-138.

45. Higgins, J.J., Rosen, D.R., Loveless, J.M., Clyman, J.C., and Grau, M.J. 2000. A gene for nonsyndromic mental retardation maps to chromosome 3p25-pter. *Neurology* **55**:335-340.

46. Yoshihara, Y., Kawasaki, M., Tamada, A., Nagata, S., Kagamiyama, H., and Mori, K. 1995. Overlapping and differential expression of BIG-2, BIG-1, TAG-1, and F3: four members of an axon-associated cell adhesion molecule subgroup of the immunoglobulin superfamily. *J. Neurobiol.* **28**:51-69.

47. Ogawa, J., Kaneko, H., Masuda, T., Nagata, S., Hosoya, H., and Watanabe, K. 1996. Novel neural adhesion molecules in the Contactin/F3 subgroup of the immunoglobulin superfamily: isolation and characterization of cDNAs from rat brain. *Neurosci. Lett.* **218**:173-176.

48. Saito, H., Mimmack, M., Kishimoto, J., Keverne, E.B., and Emson, P.C. 1998. Expression of olfactory receptors, G-proteins and AxCAMs during the development and maturation of olfactory sensory neurons in the mouse. *Brain Res. Dev. Brain Res.* **110**:69-81.

49. Kamei, Y., Takeda, Y., Teramoto, K., Tsutsumi, O., Taketani, Y., and Watanabe, K. 2000. Human NB-2 of the contactin subgroup molecules: chromosomal localization of the gene (CNTN5) and distinct expression pattern from other subgroup members. *Genomics* **69**:113-119.

50. Murai, K.K., Misner, D., and Ranscht, B. 2002. Contactin supports synaptic plasticity associated with hippocampal long-term depression but not potentiation. *Curr. Biol.* **12**:181-190.

51. Berglund, E.O., Murai, K.K., Fredette, B., Sekerkova, G., Marturano, B., Weber, L., Mugnaini, E., and Ranscht, B. 1999. Ataxia and abnormal cerebellar microorganization in mice with ablated contactin gene expression. *Neuron* **24**:739-750.

52. Montag-Sallaz, M., Schachner, M., and Montag, D. 2002. Misguided axonal projections, neural cell adhesion molecule 180 mRNA upregulation, and altered behavior in mice deficient for the close homolog of L1. *Mol. Cell. Biol.* **22**:7967-7981.

53. Montag-Sallaz, M., Baarke, A., and Montag, D. 2003. Aberrant neuronal connectivity in CHL1-deficient mice is associated with altered information processing-related immediate early gene expression. *J. Neurobiol.* **57**:67-80.

54. Rolf, B., Lang, D., Hillenbrand, R., Richter, M., Schachner, M., and Bartsch, U. 2003. Altered expression of CHL1 by glial cells in response to optic nerve injury and intravitreal application of fibroblast growth factor-2. *J. Neurosci. Res.* **71**:835-843.

55. Zeng, L., Zhang, C., Xu, J., Ye, X., Wu, Q., Dai, J., Ji, C., Gu, S., Xie, Y., and Mao, Y. 2002. A novel splice variant of the cell adhesion molecule contactin 4 (CNTN4) is mainly expressed in human brain. *J. Hum. Genet.* **47**:497-499.

56. Frints, S.G., Marynen, P., Hartmann, D., Fryns, J.P., Steyaert, J., Schachner, M., Rolf, B., Craessaerts, K., Snellinx, A., and Hollanders, K. et al. 2003. CALL interrupted in a patient with non-specific mental retardation: gene dosage-dependent alteration of murine brain development and behavior. *Hum. Mol. Genet.* **12**:1463-1474.

57. State, M.W., Greally, J.M., Cuker, A., Bowers, P.N., Henegariu, O., Morgan, T.M., Gunel, M., DiLuna, M., King, R.A., and Nelson, C. et al. 2003. Epigenetic abnormalities associated with a chromosome 18(q21-q22) inversion and a Gilles de la Tourette syndrome phenotype. *Proc. Natl. Acad. Sci. U. S. A.* **100**:4684-4689.

58. Stephens, J.C., Schneider, J.A., Tanguay, D.A., Choi, J., Acharya, T., Stanley, S.E., Jiang, R., Messer, C.J., Chew, A., and Han, J.H. et al. 2001. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**:489-493.

59. Kruglyak, L., and Nickerson, D.A. 2001. Variation is the spice of life. *Nat. Genet.* **27**:234-236.

60. Grimbacher, B., Dutra, A.S., Holland, S.M., Fischer, R.E., Pao, M., Gallin, J.I., and Puck, J.M. 1999. Analphoid marker chromosome in a patient with hyper-IgE syndrome, autism, and mild mental retardation. *Genet. Med.* **1**:213-218.

61. Nasr, A., and Roy, M. 2000. Association of a balanced chromosomal translocation (4; 12)(q21.3; q15), affective disorder and autism. *J. Intellect. Disabil. Res.* **44 (Pt 2)**:170-174.

62. Ramanathan, S., Woodroffe, A., Flodman, P.L., Mays, L.Z., Hanouni, M., Modahl, C.B., Steinberg-Epstein, R., Bocian, M.E., Spence, M.A., and Smith, M. 2004. A case of autism with an interstitial deletion on 4q leading to hemizygosity for genes encoding for glutamine and glycine neurotransmitter receptor sub-units (AMPA 2, GLRA3, GLRB) and neuropeptide receptors NPY1R, NPY5R. *BMC Med. Genet.* **5**:10.

63. Herault, J., Perrot, A., Barthelemy, C., Buchler, M., Cherpi, C., Leboyer, M., Sauvage, D., Lelord, G., Mallet, J., and Muh, J.P. 1993. Possible association of c-Harvey-Ras-1 (HRAS-1) marker with autism. *Psychiatry Res.* **46**:261-267.

64. Herault, J., Martineau, J., Petit, E., Perrot, A., Sauvage, D., Barthelemy, C., Mallet, J., Muh, J.P., and Lelord, G. 1994. Genetic markers in autism: association study on short arm of chromosome 11. *J. Autism Dev. Disord.* **24**:233-236.

65. Herault, J., Petit, E., Martineau, J., Perrot, A., Lenoir, P., Cherpi, C., Barthelemy, C., Sauvage, D., Mallet, J., and Muh, J.P. 1995. Autism and genetics: clinical approach and association study with two markers of HRAS gene. *Am. J. Med. Genet.* **60**:276-281.

66. Comings, D.E., Wu, S., Chiu, C., Muhleman, D., and Sverd, J. 1996. Studies of the c-Harvey-Ras gene in psychiatric disorders. *Psychiatry Res.* **63**:25-32.

67. Sweatt, J.D., Weeber, E.J., and Lombroso, P.J. 2003. Genetics of childhood disorders: LI. Learning and memory, Part 4: Human cognitive disorders and the ras/ERK/CREB pathway. *J. Am. Acad. Child Adolesc. Psychiatry* **42**:741-744.

68. Weeber, E.J., Levenson, J.M., and Sweatt, J.D. 2002. Molecular genetics of human cognition. *Mol. Interv.* **2**:376-91, 339.

69. Costa, R.M., Federov, N.B., Kogan, J.H., Murphy, G.G., Stern, J., Ohno, M., Kucherlapati, R., Jacks, T., and Silva, A.J. 2002. Mechanism for the learning deficits in a mouse model of neurofibromatosis type 1. *Nature* **415**:526-530.

70. Sweatt, J.D., and Weeber, E.J. 2003. Genetics of childhood disorders: LII. Learning and memory, part 5: human cognitive disorders and the ras/ERK/CREB pathway. *J. Am. Acad. Child Adolesc. Psychiatry* **42**:873-876.

71. Yamagata, T., Aradhya, S., Mori, M., Inoue, K., Momoi, M.Y., and Nelson, D.L. 2002. The human secretin gene: fine structure in 11p15.5 and sequence variation in patients with autism. *Genomics* **80**:185-194.

72. Kleinjan, D.J., and van Heyningen, V. 1998. Position effect in human genetic disease. *Hum. Mol. Genet.* **7**:1611-1618.

73. Perkins, D.O., Jeffries, C., and Sullivan, P. 2005. Expanding the 'central dogma': the regulatory role of nonprotein coding genes and implications for the genetic liability to schizophrenia. *Mol. Psychiatry* **10**:69-78.

74. Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**:185-188.

75. Wan, M., Lee, S.S., Zhang, X., Houwink-Manville, I., Song, H.R., Amir, R.E., Budden, S., Naidu, S., Pereira, J.L., and Lo, I.F. et al. 1999. Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. *Am. J. Hum. Genet.* **65**:1520-1529.

76. Cook, E.H.,Jr, Lindgren, V., Leventhal, B.L., Courchesne, R., Lincoln, A., Shulman, C., Lord, C., and Courchesne, E. 1997. Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am. J. Hum. Genet.* **60**:928-934.

77. Schroer, R.J., Phelan, M.C., Michaelis, R.C., Crawford, E.C., Skinner, S.A., Cuccaro, M., Simensen, R.J., Bishop, J., Skinner, C., and Fender, D. et al. 1998. Autism and maternally derived aberrations of chromosome 15q. *Am. J. Med. Genet.* **76**:327-336.

HARVEY CUSHING/JOHN HAY WHITNEY MEDICAL LIBRARY

MANUSCRIPT THESES

Unpublished theses submitted for the Master's and Doctor's degrees and deposited in the Medical Library are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but passages must not be copied without permission of the authors, and without proper credit being given in subsequent written or published work.

This thesis by
has been used by the following person, whose signatures attest their acceptance of the above restrictions.

NAME AND ADDRESS **DATE**